

Acknowledgements

I wish to acknowledge an enormous debt to Michael Halliday for his detailed criticisms and helpful advice during the preparation of this book. To Rodney Huddleston I am also grateful for prompting my interest in the topic and for his interest in the project.

Finally to my wife Kathy, to whom this book is dedicated, especial thanks are due for her patient support and cheerful handling of the typing. I thank her, and our children, Scott, Lisa and Belinda, for putting up with the sacrifices that the preparation of the book has necessitated.

1 Approaches to the study of cleft and pseudo-cleft constructions

1.1 INTRODUCTION

This book examines the communicative properties of 'cleft' and 'pseudo-cleft' constructions in contemporary English. These properties, I shall suggest, cannot be ignored in any attempt to provide an adequate grammatical description of the constructions. Furthermore they provide a source of explanations for the patterns of stylistic variation displayed by clefts and pseudo-clefts.

The book reports findings from a corpus-based study of clefts and pseudo-clefts in modern British English. As well as providing information on frequencies of occurrence across a range of genres, the corpus-based approach has the salutary effect of requiring that attention be paid to 'untidy' data of a type often overlooked, or ignored, in studies based on introspectively derived examples.

In this chapter the approach adopted in the study is discussed, and the assumptions upon which it is based are explored.

1.2 WHAT ARE CLEFTS AND PSEUDO-CLEFTS?

The first step is to introduce the constructions under review. Prototypical examples are presented below. Corresponding to the 'simple' or 'non-cleft' sentence in (1a), there are in English sentences of the type (1b) and (1c) (commonly referred to in formal grammars as 'pseudo-clefts' and 'clefts' respectively).

- (1) a. Tom offered Sue a sherry. *copula*
b. What Tom offered Sue was a sherry. *pseudo-cleft*
c. It was a sherry that Tom offered Sue. *cleft*

By contrast with (1a), in both (1b) and (1c) material is divided into (two distinct sections) assigned to different clauses. The part immediately following the copula within the superordinate clause (i.e. *a sherry* in both (1b) and (1c)), which normally consists of or contains

2 Cleft and pseudo-cleft constructions in English

a stressed item, is often referred to in the literature as the 'focus'. In order to avoid any confusion with Halliday's use of the term,¹ I shall refer to this constituent informally as the 'highlighted element'.² The constituent introduced by the relative pronoun, often referred to as the 'presupposition', I shall refer to as the 'relative clause'.

(Clefts and pseudo-clefts are identifying constructions, expressing a relationship of identity between the elements realized as the highlighted element and the relative clause.) In both (1b) and (1c) the highlighted element, *a sherry*, is identified as the thing which satisfies the definition provided in the relative clause, *what/that Tom offered Sue*. As identifying constructions, clefts and pseudo-clefts need to be distinguished from superficially similar attributive constructions.³ A more extensive discussion of the identifying-attributive distinction is undertaken in Section 3.3 below. Whereas identification is a relationship between an entity and some attribute that is ascribed to it, be it an indication of class membership, a quality, rule or other such characteristic, identification is a relationship between two entities, the one serving to define the identity of the other. While identifying constructions are typically reversible, attributive constructions are not (a difference explained by Halliday (1968:191, 1985:114) in terms of voice: in identifying clauses there are two 'participants' and thus both an active and passive, each with a different participant as subject: in attributive clauses there is only one 'participant', and thus only one element in the clause to function as subject). Compare (1a) with the attributive sentence in (2) below:

(2) What Tom offered Sue was too sweet.

(Whereas (1b) can be reversed (*A sherry was what Tom offered Sue*), (2) cannot be (**Too sweet was what Tom offered Sue*).⁴)

(Clefts are an exception to the generalization that identifying constructions are reversible. Here the structural device of predication, whereby the identifier is presented as complement to the non-referential subject *it*, prevents the possibility of reversibility. Like pseudo-clefts, clefts need to be distinguished from superficially similar-attributive constructions. Consider (3):)

(3) (What caused that stain on the carpet?)
It was a sherry that Tom offered Sue.

(Unlike a cleft, the sentence in (3) has a genuine anaphoric pronoun (*it*) as subject, and a *that*-clause as dependent relative within the noun

Approaches to cleft and pseudo-cleft constructions 3

phrase headed by *sherry*. Whereas (1a) is a paraphrase of the cleft in (1c), it is not a paraphrase of the sentence in (3), and would be quite inappropriate in the context. It may further be noted that when spoken, attributive sentences like that in (3) will differ from (intonationally unmarked) clefts in that, other things being equal, they will carry a nuclear stress on an item in the relative clause rather than on the head of the noun phrase of which it is a modifier.)

(The primary function of 'pseudo-clefts' and 'clefts', as the names suggest, is thematic: they enable subsets of elements to be grouped into two parts in an almost unlimited number of ways (cp. *What Tom did was offer Sue a sherry*; *The one who offered Sue a sherry was Tom*; *It was Tom who offered Sue a sherry*). Furthermore, as noted above, (the highlighted element and relative clause of pseudo-clefts may be inverted (e.g. *Tom was the one who offered Sue a sherry*). These constructions (pseudo-clefts with highlighted element as theme) are referred to in this book as 'reversed pseudo-clefts', while their non-reversed counterparts are termed 'basic pseudo-clefts'. The thematic properties of clefts and pseudo-clefts are discussed in detail in Chapter 5.)

1.3 A FUNCTIONAL PERSPECTIVE

One may identify two major traditions in the grammars of the second half of this century, one formal (syntagmatic in orientation, with roots in logic and philosophy, and represented in its most influential form by Chomsky's transformational-generative model), the other functional (often paradigmatic in orientation, with roots in ethnography and rhetoric, and perhaps best represented by Halliday's systemic-functional model). The influence of this opposition is evident in recent studies of clefts and pseudo-clefts. A number of formal studies have appeared, devoted mainly to syntax and written mostly within the transformational framework.⁴ Several contributors to this tradition acknowledge a debt to Jespersen,⁵ who gave cleft sentences their name, and whose work in many ways foreshadows later accounts.⁶

Functional treatments are considerably fewer in number, and tend to be characterized by greater 'theory-neutrality' than those in the formal tradition (those that have exerted the greatest influence upon the present study are Halliday (1967a, 1985) and Prince (1978).) Halliday examines clefts and pseudo-clefts (which he labels 'predicated-theme structures' and 'identified-theme structures' respectively) as part of his discussion of the textual organization of the

4 Cleft and pseudo-cleft constructions in English

English clause. This discussion, in turn, is to be located within the broader context of Halliday's investigations into the text-creating component of language (termed 'functional sentence perspective' by linguists of the Prague School).

Prince's (1978) text-based study draws on the work of Chafe (1976), Grice (1975), and others in order to analyse and subclassify clefts and pseudo-clefts (which she labels 'it-clefts' and 'WH-clefts' respectively) according to their discourse functions. Whereas Halliday in his work on clefts and pseudo-clefts concentrates more on their thematic than their informational structure, Prince's concern is solely with the latter.

(This book leans more towards the functional than the formal approach.) It is assumed that, as Halliday (1970c:323) argues:

the structure of language as a whole has been built up in such a way that it reflects the demands that are made on language and the functions it is required to serve.

N Accordingly, in the case of clefts and pseudo-clefts, we might expect their structural characteristics to be reflected in the uses to which they are put in discourse. The key to why they have evolved as a resource in the English language, I shall assume, is a consideration of speakers' communicative needs – the need to select, at a given point in a discourse, a form that is appropriate in the light of assumptions about what information an addressee already possesses, the need to select a form which appropriately emphasizes or focuses upon particular parts of the message, the need to draw a certain contrast, and so forth.)

In terms of their communicative *raison d'être*, clefts and pseudo-clefts belong to a set of constructions each of which represents a marked variant of a communicatively 'simpler' sentence. Consider the following pairs:

- (4) a. Tom poured the sherry.
b. The sherry was poured by Tom.
- (5) a. Tom offered Sue a sherry.
b. Tom offered a sherry to Sue.
- (6) a. That Tom offered Sue a sherry surprised no one.
b. It surprised no one that Tom offered Sue a sherry.

The passive clause in (4b) retains the propositional meaning of the active clause in (4a), but differs from it in terms of communicative significance. Whereas (4a) is 'about' Tom (the theme, or topic), (4b)

Approaches to cleft and pseudo-cleft constructions 5

is about the sherry. Furthermore, assuming a neutral intonation, (4a) and (4b) will reflect different divisions between 'given' and 'new' material. Thus in a context where it is given (that is, already established or able to be taken for granted) that Tom poured something, and new that the something was a sherry, (4a) is likely to be preferred. Conversely, if it is given that the sherry was poured by someone, and new that that person was Tom, then (4b) is likely to be preferred. An additional communicative feature of the passive construction is the omissibility of the agent (*The sherry was poured*), by contrast with the non-omissibility of its active analogue, the subject (**Poured the sherry*). Agentless passives are favoured in contexts where the identity of the agent is unknown, readily inferable, unable to be revealed, or the like.

The difference between (5a) and (5b) similarly involves communicative rather than propositional dimensions of meaning. (5a) is likely to be used in a context where a sherry represents new information, and (5b) where Sue is new.

(6a) differs from (6b) in the extraposition of the subordinate *that*-clause to the right of *surprised no one*, and the appearance of *it* in subject position. The main communicative effect of this 'transformation' is to place a 'heavy', or grammatically complex, unit at the end of the sentence to facilitate processing by the addressee. This effect is relevant to questions of information (a heavy unit will, other things being equal, be more likely than a light one to contain new information, and thus will tend to be preferred later rather than earlier within the sentence.)

Like the constructions in (4b), (5b) and (6b), clefts and pseudo-clefts represent communicative variants of structurally simpler sentences, differing from them in terms of theme-selection, the presentation of given and new information, and the like. Work on the organization of language as an instrument of communication, using notions such as theme and information, has not figured prominently in linguistic theory. The first impetus to the study of this aspect of language came from members of the Prague School,⁷ who refer to it as 'functional sentence perspective'. This form of linguistic organization was incorporated by Halliday into his theory of systemic grammar (see especially Halliday 1967a–68) under the general heading of 'theme', in a tripartite division of grammatical organization in which the other major components were 'transitivity' and 'mood'. These three major components are derived in the theory from three major functions of language: 'transitivity' from the 'experiential' (or 'ideational') function, which encodes our experience of the world and is

concerned with the expression of content; 'mood' from the 'interpersonal' function, which is concerned with the expression of social and personal relations including all forms of the speaker's intrusion into the speech situation; and 'theme' from the 'textual' function, which 'fills the requirement that language should be operationally relevant – that it should have a texture, in real contexts of situation, that distinguishes a living message from a mere entry in a grammar or dictionary' (1973:42). One of Halliday's important achievements was to elaborate and distinguish, within his treatment of the textual dimension of language, the concepts of information structure and thematization. Information, or information structure, the organization of discourse into a linear succession of units, each internally structured in terms of 'given' and 'new' components of the message, is claimed to be signalled phonologically by intonation. As we shall see, not everyone agrees with Halliday as to what phonological variables expound the system of information, and many would accord more weight to other kinds of variables – syntactic and contextual – particularly in attempting to analyse the information structure of written texts. Thematization, or theme–rheme structure, is the name used for the division of the clause into two sections, determined, in English as in most languages, by the sequential ordering of sentence elements: the theme is represented by the element(s) assigned to initial position, and the rheme is the remainder of the message.

1.4 REGISTER

In this book attention is paid to broad stylistic considerations – patterns of distribution across different language registers – of a type not commonly regarded as falling within the province of a grammatical study. What is the justification for this? In order to answer this question it is necessary for me to spell out certain assumptions being made here regarding the relationship between language and context. We have seen that the functional approach that I am adopting is a comprehensive one, in its attempt to explain the nature of grammar in functional terms. In fact the framework is even broader than this. It provides principled means of relating the situational context to the types of discourse that can be predicted to occur within them. The term 'register' is widely used to refer to a variety of language distinguished according to the use to which it is put (as opposed to a variety characterized in terms of the properties of its users, a 'dialect'). Ure and Ellis (1977:197) define register as 'a certain kind of language patterning regularly used in a certain kind of situation. It is

a social convention. By making use of different kinds of register patterns, people show that they are aware of the social situations in which they find themselves.' (Halliday (1977:201, 1978:21, etc.) regards situation as a semiotic construct comprising three dimensions, 'field', 'tenor' and 'mode'. Field is the ongoing social action, including the 'subject matter' of the discourse, tenor concerns the role relationships between interactants, and mode concerns the role that language plays, including the channel or medium of communication. Choices made in terms of field, tenor, and mode will bias the selection of linguistic options. Specifically, Halliday (1975:36) proposes that each situational component will be reflected in the types of choice made from one particular functional component of the linguistic system:

the type of symbolic activity (field) tends to determine the range of meaning as content, language in the observer function (ideational); the role relationships (tenor) tend to determine the range of meaning as participation, language in the intruder function (interpersonal); and the rhetorical channel (mode) tends to determine the range of meaning as texture, language in its relevance to the environment (textual).

What Halliday is suggesting is that if we compare two texts differing in field (say, a scientific paper on physics and one on chemistry) the most likely differences will be concerned with the expression of processes, participants, circumstances, and the like. (If we compare two texts differing in tenor (say, a conversation with a friend about a recent event, and one with a stranger at a bus stop), it is more likely that the differences will be in the areas of styles of address, expressions of attitude and other such interpersonal features. And finally, if we compare two texts differing in mode (say, a discussion about a film with a professor and an essay on the film written for the professor) the differences are likely concern theme, information structure and various cohesive devices.)

In Chapter 2 Halliday's hypothesis concerning the interdependence of text-type and situation-type is explored in more detail, and a discussion of the register dimensions of the database used for the investigation of clefts and pseudo-clefts is presented. Not unexpectedly, given the relevance to the description of clefts and pseudo-clefts of textual aspects of meaning, the most revealing register comparisons were found to be those involving text-types differing in mode. The choice of a written and spoken corpus was in fact made precisely in order to facilitate such comparisons.

The preliminary consideration in this section of the interface between language structure and the structure of social context is intended to provide an appropriate perspective for the discussion in Chapter 6 of the distributional patterns displayed by clefts and pseudo-clefts across a range of text-types in the two corpora. The present discussion is also relevant to the taxonomy which is proposed in Chapter 5 to account for the categories of given information associated with pseudo-clefts. It is argued that the sources from which such information may be presented as recoverable by the speaker are either co-textual or contextual. In the latter case there are three types of source available, corresponding to the environmental determinants of text, namely field, tenor and mode.

1.5 STRUCTURE OF THE BOOK

In Chapter 2 the advantages of the corpus-based approach are discussed, and the nature and composition of the database are described.

In Chapter 3 an attempt is made to define and delimit the class of clefts and pseudo-clefts in English in a theoretically defensible manner. Three subclasses of pseudo-clefts are proposed: *wh*-clefts, *it*-clefts, and *all*-clefts. It is claimed that most of the difficulties involved in separating clefts and pseudo-clefts from superficially similar constructions are associated with the distinction between identification and attribution.)

Chapter 4 focuses on the formal properties of clefts and pseudo-clefts. A number of the syntactic analyses that have been proposed are examined and their adequacy assessed. Particular attention is paid to the range of classes of highlighted elements that are accepted by clefts and pseudo-clefts, and to the functions that they occupy in the corresponding non-cleft version.

Chapter 5 reviews the literature on the concepts of givenness and newness, and theme and rheme, and establishes the components of a description of clefts and pseudo-clefts in terms of these textual functions.)

In Chapter 6 an informational/thematic description of clefts and pseudo-clefts in the database is presented, along with relevant statistics from the database. It is established that there are substantial communicative differences between the constructions: pseudo-clefts are oriented towards givenness and generate thematic prominence of an ideational nature; clefts are oriented towards newness and generate thematic prominence of a textual nature.)

In Chapter 7 the distribution of cleft and pseudo-cleft constructions across a range of register categories is investigated, and explanations for the distributional differences that emerge are advanced which take into account properties of the constructions noted in previous chapters.

Finally, in a short eighth chapter, a summary of the findings and claims of the study is presented.

2 The database

2.1 A CORPUS-BASED APPROACH

Despite widespread acknowledgement among linguists in recent years that such discourse-relevant notions as information, theme, and topic must be invoked in explaining the structural and functional characteristics of cleft and pseudo-cleft constructions,¹ few text-based studies have been conducted to date,² and no corpus-based studies. Consequently our knowledge of the discourse functions and register distribution of these constructions has remained meagre and lacked a firm empirical base.

There has been a long tradition of text-based study in English grammars. However text-based study is to be distinguished from corpus-based study. Svartvik and Quirk (1980:9-10) observe that while most of the statements in the earlier comprehensive grammars of scholars such as Poutsma, Kruisinga and Jespersen had textual support, their approach differs in at least four ways from a corpus-based one:

- (a) no single text was subjected to complete description – the fundamental 'total accountability' principle of corpus study; (b) no attempt was made to draw even the eclectic examples from texts representative of the whole range of usage (there was a heavy bias in Jespersen, for example, for *printed* material, especially *fiction*); (c) it is not always easy to tell whether a textual sample from the 18th century is valid for 20th-century usage (or even whether the grammarian meant us to believe it so); and, most obviously, (d) it is only in the second half of the twentieth century that it has become practicable to record *spoken* material and hence to subject

to properly detailed study the most natural form of language and the one that is overwhelmingly dominant for any individual, in whatever walk of life.

It is now widely recognized in linguistic research that corpora are useful, if not necessary, for some forms of linguistic research. Using a corpus, the linguist can make more objective statements than introspective analysis permits, and state frequencies in a variety of uses of the language. No individual can be expected to have an adequate command of the whole grammatical repertoire of a language, and no grammarian can be expected to describe adequately the grammatical properties of the whole repertoire from his or her own unsupplemented resources. Even when linguists are working in those areas of the repertoire in which they are most likely to be thoroughly competent (for example informal conversation) it is difficult to guard against idiolectal bias, accidental omissions and distortions, in the absence of objective data.

The present study is the first corpus-based investigation of cleft and pseudo-cleft constructions in English to be undertaken. Every such construction – as here defined, and totalling 1,785 tokens – was analysed in two standard computer corpora of present-day British English: the London-Lund Corpus, representing spoken language, and the Lancaster-Oslo/Bergen Corpus, representing written language (together referred to in this study as 'the corpus'). This choice of corpora (the only other readily available large computer-corpora of present-day English being the Brown University Corpus of American English) enabled comparisons between speech and writing to be made, while dialect was held constant. Some of the findings of the study are, then, offered as a contribution to our understanding of the differences between speech and writing (see Chapter 7).

While there are many benefits in a corpus-based approach, it also has certain limitations. No corpus, however large, can be expected to give information in the requisite degree of detail on all the grammatical structures of English. Even extremely large samples of the language will fail to contain constructions which, though statistically infrequent, are readily accepted by native speakers as grammatical. It was this consideration which led Randolph Quirk and his co-workers in the Survey of English Usage to develop the notion of the 'extended corpus', a random sample which is extended in selected directions by direct elicitation techniques. It has occasionally been found necessary in this study to resort to invented examples (tested against the intuitions of acquaintances if their acceptability was in question) in

12 Cleft and pseudo-cleft constructions in English

order to provide illustrative data where none was available from the two corpora used.

2.2 THE NATURE AND COMPOSITION OF THE DATABASE

A description of the London-Lund Corpus (henceforth 'LL') must begin with the project of which it was a later development, the Survey of English Usage. The aim of the Survey, based at University College London from 1960, was to describe the grammatical repertoire of adult educated speakers of British English on the basis of information derived from a corpus of 'text' material and from elicitation tests. The features of this target variety, 'educated English', are described by Quirk (1974:168) as follows:

شخص
This is a shorthand characteristic of the people who write books, who lead our cultural, educational, political, and religious life, who constitute the major centre of imitation in language, in the sense that it is these people's English that is thought to be best by society as a whole, these people's English that is taught to foreign learners, and so on.

شخص
It was decided to compile a million-word corpus consisting of 200 'texts', each of 5,000 running words, produced in the period since 1950. Half of the texts represent material with origin in writing (46 printed, 36 non-printed, and 18 'as spoken'), and half represent material with origin in speech (24 monologue and 76 dialogue). The allotting of proportions of the corpus was not performed with statistical accuracy in mind. Instead, the criterion used was the amount of data needed to represent adequately a variety's grammatical/stylistic potential.

A sister project of the Survey of English Usage, the Survey of Spoken English, was set up at Lund University in 1975 with the primary aim of converting the spoken material (or, more accurately, 'material with origin in speech') of the London project into machine-readable form. The computerized version, LL, consists of 87 texts (totalling 435,000 words). Information about the 87 texts, presented on p.1 of the magnetic computer tape, is reproduced in Table 2.1.

As Table 2.1 shows, subclassification within the category of conversation results from a division into conversation between intimates, between equals, and between disparate participants. Further, conversation is divided as between business and social matters. Cutting across this categorization is another whereby all texts (conversation,

Table 2.1 Text description of the London-Lund Corpus

Text categories	No. of texts	Approx. no. of words	Description
S.1-S.3	34	170,000	Spontaneous, surreptitiously recorded conversations between intimates and distant.
S.4	7	35,000	Conversations between intimates and equals. These texts are either wholly non-surreptitious or composite, i.e. individual speakers can be marked for surreptitiousness.
S.5	11	55,000	S.5.1-7: Non-surreptitious public conversations between equals. S.5.8-11: Non-surreptitious private conversations between equals.
S.6	6	30,000	Non-surreptitious conversations between dispartes. All texts except S.6.2 are interviews or conversations.
S.7	3	15,000	Surreptitious telephone conversations between personal friends.
S.8	4	20,000	Surreptitious telephone conversations between business associates.
S.9	3	15,000	Surreptitious telephone conversations between dispartes.
S.10	8	40,000	Spontaneous commentary. S.10.1-4: Sport (cricket, football, boxing, horse-racing). S.10.5-8: Non-sport (state funeral, royal wedding, physics demonstration, etc.).
S.11	5	25,000	Spontaneous oration (case in court, dinner speech, radio 'My word', recordings in the House of Commons).
S.12	6	30,000	Prepared but unscripted oration (sermons, university lectures, cases in court, political speech, popular lecture)
Total	87	435,000	

14 Cleft and pseudo-cleft constructions in English

prepared oration, spontaneous oration and spontaneous commentary) are classified according to features relating to the speech situation. These are described by Svartvik and Eeg-Olofsson (1982:88) as follows:

DIALOGUE with two or more participants in contrast to monologue with only one speaker, which is typical of, for example, such activities as commentaries or demonstrations.

FACE-TO-FACE, where gestures and other forms of non-verbal communication clearly play an important part, as opposed to other types of interaction, such as telephone conversation and radio commentary where this is not the case.

PRIVATE, i.e. not recorded before an audience or on radio, e.g. a relaxed, informal conversation between friends, as opposed to public performance, including radio debates, interviews, commentaries and audio-conditioned orations.

SURREPTITIOUS recording, where the participants are not aware of being recorded, as opposed to an 'open' recording with the microphone visible.

RADIO broadcast as compared with other, non-public, recording occasions, such as telephone or face-to-face conversations.

The 87 texts in LL are classified according to these five situational factors in the matrix of Table 2.2 (adapted from Svartvik and Eeg-Olofsson 1982:86).

A prosodic transcription that was evolved in the Survey, while allowing orthographic (non-phonetic) representation wherever possible, was sensitive to a wide range of prosodic and paralinguistic features.³ In LL the number of features, particularly paralinguistic, was reduced. Omitted features include tempo, loudness, and modifications in voice quality; retained are such basic distinctions as tone unit, nucleus, booster, onset, and stress.⁴ In the present study it was found that even this reduced set of features provided more information than was necessary in the investigation of informational, thematic, and other properties of clefts and pseudo-clefts. For this reason the number of features has been decreased even further, to speaker details (speaker identity, contextual comments on non-linguistic activities such as laughing and coughing, and incomprehensible words), tone unit (the basic prosodic unit of analysis), nucleus (the peak of greatest prominence in the tone unit, signalled by a change of pitch), and pauses of various length. A complete list of

Table 2.2 Text classification of the London-Lund Corpus

Description of text groups	Sub-groups	Dialogue	Face-to-face	Private	Surreptitious	Radio	Text labels	Sub-group total	Text group total
Face-to-face conversation	a	+	+	+	+	-	S.1, S.2, S.3	34	46
	b	+	+	+	-	-	S.4, S.5, S.8-11, S.6.2	12	
Telephone conversation	c	+	-	+	+	-	S.7, S.8, S.9, S.10	10	10
Discussion, interview, debate	d	+	+	-	-	+	S.5, 1-7, S.6.1, S.6.3-5, S.10.8	12	12
Public, non-prepared commentary, demonstration, oration	e	+	+	-	-	-	S.11.1, S.11.4-5	3	
	f	-	+	-	-	+	S.6.6, S.11.3	2	12
	g	-	-	-	-	+	S.10.1-7	7	
Public, prepared oration	h	-	+	-	-	-	S.11.2, S.12	7	7
		71	70	56	44	21			
		No. of texts					87	87	87

symbols used is presented in the 'List of abbreviations and symbols' on pp.xii-xiii.

The Lancaster-Oslo/Bergen Corpus (henceforth 'LOB') originated in a project, begun in 1970 at the University of Lancaster, whose aim was to compile a computer corpus of printed British English designed to match the Brown University Corpus of American English. The LOB project was completed in 1978 with the assistance of Stig Johansson at Oslo and the Norwegian Computing Centre for the Humanities at Bergen.

Like its American counterpart, LOB aims at a general representation of text types, rather than concentrating upon limited types of texts to be used for specific purposes. Table 2.3 (from Johansson, Leech and Godluck 1978:3) summarizes the basic composition of the British material, as compared with the American corpus. Fifteen different genres are represented, the selection of extracts in each being close-to-random. The extracts themselves are relatively short (2,000 words) and number 500, giving a total of around one million words.

There is one caveat that must be entered regarding the composition of Brown and LOB: the absence of non-printed (that is, handwritten and typewritten) material from these corpora limits their usefulness for studies of the written language generally (see further Section 2.3 below). This limitation must be borne in mind when, in the present study (especially in Chapter 7), broad comparisons of speech and writing are made.

2.3 REGISTER CHARACTERISTICS OF THE DATABASE

The text categories of LL and LOB embrace a wide variety of configurations of the register dimensions of field, tenor, and mode. Let us consider each of these in turn.

(i) Field

The fields – institutions or social processes in which language plays a part – which are represented in LL and LOB range from those of a technical nature such as linguistics and biology, which tend to so determine the language used that in many cases it becomes rather restricted to that role and to those acquainted with it, to non-technical fields such as personal experience and imagination, which are not so restricted.⁵ To put this another way, technical fields will tend to involve linguistic realizations (especially lexical rather than grammatical) that are characterized by a high degree of probability, and in a significant number of cases by indexical realizations (that is,

Table 2.3 The basic composition of the Lancaster-Oslo/Bergen and Brown corpora

Text categories	Number of texts in each category	
	Brown	LOB
A Press: reportage	44	44
B Press: editorial	27	27
C Press: reviews	17	17
D Religion	17	17
E Skills, trades, and hobbies	36	38
F Popular lore	48	44
G Belles lettres, biography, essays	75	77
H Miscellaneous (government documents, foundation reports, industry reports, college catalogue, industry house organ)	30	30
J Learned and scientific writings	80	80
K General fiction	29	29
L Mystery and detective fiction	24	24
M Science fiction	6	6
N Adventure and western fiction	29	29
P Romance and love story	29	29
R Humour	9	9
Total	500	500

those whose use as single items is sufficient to enable an addressee familiar with the field to recognize it).

Features of field tend to be realized differently in speech and writing. The dialogic texts of LL are typified by a sense of shared fields, of mutually negotiated institutional focus. This characteristic is largely absent from the written texts of LOB. (In writing the sense of institutional focus must be achieved by the writer. The reader, who does not contribute to the negotiation of the field, must discover it. The ease with which this is achieved depends on the reader's familiarity with the writer's institutional framework.) There is one significant exception to this field difference between the dialogic texts of LL and the texts of LOB: namely, the dialogue that occurs in the fictional texts (in which the distribution of cleft and pseudo-clefts is, as one might expect, more like that in spoken than written genres).

The most strongly represented genre in LL, conversation (which is represented – with the exception of the interviews and debates in S.6

— by text categories S.1–9), encompasses a wide range of fields. By contrast, text categories S.10–S.12, representing genres such as radio commentary, sermons and lectures, comprise a range of fields that is not only narrower but also typically more technical. These differences are evidently closely related to the tenor and mode differences between the genres.)

Field appears to play a more significant role in determining the text categories of LOB than of LL. This is particularly the case with the subclassification of 'Imaginative Prose' into 'General fiction', 'Mystery and detective fiction', 'Science fiction', 'Adventure and western fiction', 'Romance and love story', and 'Humour'. With the non-fictional categories, mode and tenor differences exert slightly more influence, but again it is one that is systematically linked to that of field. Press reportage, for instance, is characterized by relative proximity of content and of readership (see below for discussion), mode features which correspond to the current relevance of the field. With both essays and legal documents the content and the readership are distant (see below for discussion), mode features which again are field-related (in this instance, corresponding to the general applicability of the field: oriented towards universality in the case of essays, towards conditions and contingencies in the case of legal documents).

In the distributional analysis of cleft and pseudo-cleft constructions across the two corpora it was found that numbers of tokens in individual categories were often too small to be significant. For this reason text categories were organized into several broader groupings. The principal subgrouping in LOB, that between 'informative' or 'non-fictional' categories (A–J in Table 2.3), and 'imaginative' or 'fictional' categories (K–R in Table 2.3), is primarily field-based, relating mainly to the actuality or otherwise of the content of the text. The validity of this distinction is confirmed by the findings of Biber (1985, 1986) in a corpus-based study using factor analysis (see further below), showing that the fiction versus non-fiction opposition is a significant discriminator of text-types.

(ii) *Tenor*

Tenor — the relationship that the speaker has with the addressee — has been subclassified by Martin (1984a:27) into the dimensions of 'status', 'affect' and 'contact'. Status resolves itself into two dimensions, 'power' and 'solidarity'. Power relations arise where there are differences in status, based mainly on social class, age and sex, yielding configurations such as adult-child, teacher-student, and boss-employee. Solidarity is associated with sameness of status, reinforced by positive feelings of comradeship. Affect refers to the speaker's

feelings towards the addressee (on a scale ranging from love to hate). Contact refers to the frequency with which the speaker encounters the addressee (on a scale from rarely to often).

Tenor variations are reflected largely in variations in the formality of the language used. Of the various attempts that have been made to classify such variations, Joos (1967) is probably the best known. Each of his five 'styles' ('frozen', 'formal', 'consultative', 'casual', and 'intimate') is described in terms of a range of text-types.

Variations of tenor are considerably smaller in LOB than in LL. The texts of LOB, being all of printed material, are relatively uniform in the level of formality of the language used. This is not to deny that differences exist (for example press reportage, where the author-reader relationship tends to be more regionally and temporally circumscribed than that for learned treatises, is correspondingly less formal).

A considerable range of relationships between speakers and their addressees is represented in LL. As was observed in Section 2.2 above, there are conversations between 'intimates' (that is, personal friends), 'distant' (that is, strangers), 'equals' (for example, business associates), and 'disparates' (for example, lecturer-student). Unfortunately, LL has not been systematically subclassified by its compilers into these tenor-related categories. So, for example, text categories S.1–S.3, which represent the bulk of the face-to-face conversation, are characterized simply (and ambiguously) as 'Spontaneously, surreptitiously-recorded, conversations between intimates and distant'. The list of speakers given for each text (for example, for S.1.1, 'A male academic, age c.44; B male academic, age c.60') provides no further information on their social or personal relationships.

There is one subset of texts for which mode is constant, enabling an independent assessment of the contribution of tenor to be made. These are the telephone conversations in Text categories S.7 (conversations between personal friends), S.8 (conversations between business associates), and S.9 (conversations between disparates). For further discussion see Section 7.2 below.

(iii) *Mode*

Mode — the medium through which linguistic interaction takes place and its effects on communication — may be seen initially as a simple choice between speech and writing. However, we make use of many different media in today's society: radio, telephone, television, letters, books, and so on. The effects which each such medium has on communication are described by Martin (1984a, 1984b) in terms of

two kinds of physical distance, 'addressee-proximity' and 'content-proximity'.

Addressee-proximity has to do with how physically close the speaker is to the addressee, Martin identifies a scale ranging from face-to-face conversation, with its direct aural and visual contact, to stream of consciousness writing, where the question of audience disappears altogether. Intermediate positions are occupied by, for example, telephone, which allows aural feedback but has no visual channel, and television, which has one-way visual and aural contact. Martin's scale is outlined below.

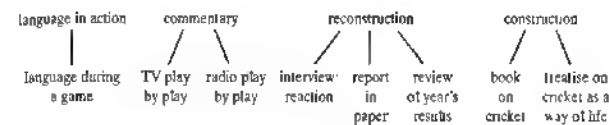
face to face	tele- phone	TV	radio	letter	book	stream of conscious- ness
+ aural + visual	+ aural - visual	one way aural and visual	one way aural - visual	delayed written feedback; - visual	review feedback; - feed- back (unless illus- trated)	audience = self

Content-proximity refers to the distance between language and what it describes. The poles of this dimension are language in action and language as reflection. Martin's (1984a:26) extended example of the various uses of language that will be more or less closely related to a game of cricket is worth quoting in full:

At one end of this scale we have the language of the players and umpire during the game. Next on the scale would be play by play commentary on the game. This will be somewhat further removed from language in action on radio than on TV since on TV the commentator and his audience can both see what is going on, but on radio only the commentator can. At a further remove from the action would be an interview with the players after the game. Then a report of the game in the paper the next day. This is leading on towards the reflective end of the scale, where action is reconstructed, rather than commented on. Next we might place a book about cricket in a given year. Then a book about cricket in general. And finally, an even more abstract text, one which constructs rather than reconstructs reality. An example might be a philosophically oriented treatise on sport, fair play, and cricket as

symbolising the English way of life (no under-arm bowling allowed). What is happening along this scale is that language is getting further and further removed from what it is actually talking about, not simply in terms of temporal distance (distance from the scene of the crime as it were), but eventually in terms of abstraction as well.

Martin outlines this action/reflection scale as shown below.



Broadly speaking, variations in mode reflect variations in the context dependency of texts. Whereas speakers engaged in mutual activity and dialogue will encode many implicit meanings, most written texts are removed from the events described, making it necessary for meanings to be made explicit.

Before discussing mode variation in the corpus I shall briefly discuss several factors which serve to differentiate speech and writing taken as broad categories: context dependence, lexical density, and grammatical intricacy. Spoken and written language are not of course separate, discrete phenomena, but manifestations of the same underlying system. Thus any postulated differences are to be regarded as general tendencies rather than invariable properties of the spoken or written medium. Indeed, certain kinds of premeditated speech, in which the speaker is either reading aloud from a written text or at least consciously structuring and monitoring what s/he says, as in debates and formal lectures, may closely resemble written language. Unspontaneous speech of the type normally used on the stage or screen has been styled 'spoken prose' by Abercrombie (1963), who explores the ways in which it differs from conversation. The most striking differences, he contends, are probably phonetic. They include the even tempo of spoken prose, compared with the frequent variations in tempo of conversation; the close relationship between pauses in spoken prose, but not conversation; and the occurrence of long periods of silence in conversation, but not spoken prose.

Speech and writing typically differ in their degrees of context dependence. The high degree of interdependence between 'typical' speech (say, casual conversation) and the situation in which it is produced, is reflected in the incidence of words and expressions whose interpretation requires knowledge of when, where, and by whom the text was produced. In a written text the resolution of such context-dependent elements is generally contingent upon information supplied within the text itself. Another consequence of the less immediate nature of written language is, as noted by Abercrombie (1971:43), that it tends to omit features of intonation and rhythm that express this involvement with the context of situation:

the whole object of written language is to be *free* of any immediate context, whether personal or situational, and that is why it dispenses with systematic indication of intonation and rhythm, only giving the vaguest of hints in the form of question marks, commas and so on.

Despite this absence of suprasegmental phenomena, different readers tend to give similar renditions of the same passage. The principle that appears to operate here is a straightforward one: if there is no linguistic indication to the contrary it will be assumed that the intonation is unmarked (that is, that the nucleus falls on the last lexical item in the clause). It is argued below that cleft constructions regularly provide just such an 'indication to the contrary' and that this provides one form of explanation for their popularity – relative to pseudo-clefts – in writing.

Lexical density refers to the proportion of content words to the total discourse. Ure (1971) demonstrates that the lexical density of a text correlates with its position on a register spectrum running from 'active' (that is, 'language-in-action', which is the least dense) to 'reflective' (which is the most dense). Since written language is typically reflective rather than active, it follows that written texts will have higher lexical density, and that this will decrease as the text moves towards spontaneous speech.

Grammatical intricacy is a more controversial type of complexity that differs across written and spoken texts. While it has often been claimed that writing is grammatically more complex than speech,⁶ Halliday (1979, 1987) has demonstrated, through analyses of natural spoken texts, that as far as sentence structure is concerned the reverse is actually the case. It is his view that the more unselfconscious the language, the more grammatically intricate it is likely to

become. The point is explained in terms of the different processing constraints associated with speech and writing (Halliday 1979:49):

Writing is a deliberate and, even with modern technology, a relatively slow process; the text is created as an object, and is perceived by the reader as an object – it exists. Spoken text does not exist: it happens. The text is created, and is perceived by the listener, as a process. Its reference points are constantly shifting; the speaker keeps on going, and the listener cannot pause and hold up the text for contemplation – he is carried along with it, tracking the process as it happens.

Halliday develops two metaphors to encapsulate this difference. The complexity of speech is characterized in terms of its 'choreographic' intricacy of movement (linguistically, complex sentence structures with low lexical density); the complexity of writing is characterized in terms of its 'crystalline' denseness of matter (linguistically, simple sentence structures with high lexical density).

Halliday's belief that extensive use of subordination is associated with the real-time production constraints characteristic of speech is supported by the findings of the study reported by Biber (1985, 1986). Using factor analysis, Biber identified groupings of linguistic features that co-occur with a high frequency across spoken and written texts and used them to examine relationships among text-types. Biber found subordination to be a feature of interactive (predominantly spoken) texts, and suggests that it is 'an important strategy for expressing fuller content under real-time production constraints, when there is little opportunity to elaborate content through exact lexical choices' (1986:395).

The register dimension of mode is relevant not only to broad differences between the spoken and written corpora taken as wholes, but also to finer differences between subcategories within them. It is argued in Chapter 7 that the different principles of organization that are found in speech and writing, and which are crucially related to the mode differences between them, provide the main source of explanation for the variable distribution of clefts and pseudo-clefts across LL and LOB.

Variations of mode are less extensive in LOB than in LL. Because LOB is a corpus solely of printed material, texts are relatively uniform in their context-dependency (though naturally there are certain differences in content-proximity). This uniformity is in fact a major limitation on the usefulness of LOB for studies of the differences between speech and writing. As already noted in Section 2.2,

the exclusion of non-printed texts (such as personal letters, memos, and diaries) from LOB restricts the range of modes (and tenors) to such a degree that one must exercise caution in using it as the basis for generalizations about the written language. Biber, who used LL, LOB and the Brown Corpus of written American English as his database, supplemented his written corpus data with a collection of ten professional letters representing the category – absent from LOB and Brown – of 'written interpersonal communication'. Biber's inclusion is a partial attempt to address the limitations of LOB and Brown, but one might require the inclusion of a broader range of typed, and also handwritten, texts, before it could be claimed that the database provided a satisfactory representation of the written language. Furthermore, it is important, in comparisons of speech and writing in which the written data – as in the case of LOB – includes prose fiction, to bear in mind the potential influence of dialogue (with its speech-like characteristics) upon the results.

By contrast with LOB, in LL mode is the primary determinant of text categorization. As noted in Section 2.2 above, not only is there a variety of physical channels represented (face-to-face, telephone, radio, and so forth), with their differing implications for the proximity of the addressee(s), but also texts are spread along the extent of the action/reflection scale, ranging from the 'active' demonstrations in Texts S.10.7 and S.10.8, to the 'reflective' university lectures in Text S.12.2.

In plotting the distribution of cleft and pseudo-cleft constructions in LL, the classification of texts into eight subgroups which is presented in Table 2.2 was invariably found to be too delicate. As in the case of LOB, it was found necessary to organize categories into several broader subgroupings. Whereas for LOB these subgroupings are mainly field-determined, for LL mode and tenor are the primary determinants. One subdivision is between 'private' categories (that is, those not recorded before an audience or on radio, including conversations between intimates and dispartes, both face-to-face and via telephone) and 'public' categories (that is, radio debates, interviews, commentaries, and audio-conditioned orations). A similar, but not entirely identical subdivision, is that between 'dialogue' (where there are two or more active participants, as in conversations, interviews, debates and discussions) and 'monologue' (where there is only one speaker, as in commentaries, lectures, sermons and speeches).

Once again, as was the case with the major subgrouping of text categories in LOB described above, the proposed subdivisions for LL

are lent support from the work of Biber. The private/public and dialogue/monologue distinctions map quite closely onto two 'textual dimensions' which he proposes on the basis of clusterings of linguistic features yielded by his factor analysis, and which he terms 'interactive/edited' and 'situated/abstract'. The continua defined by all four parameters set apart casual conversation and prepared oration, with interviews and spontaneous speeches occupying intermediate positions.

2.4 ACCESSING THE DATABASE

In order to facilitate the task of locating cleft and pseudo-cleft tokens within LL and LOB, a program was run to underline and bold face a set of key words critical for the identification of the constructions, as listed below:

- (i) All forms of the verb *be* (*be, is's, are're, am, was, were, been, being*)
- (ii) *it*
- (iii) The English interrogatives (*what, who, when, where, how, why*)
- (iv) The 'pro-form' equivalents of the English interrogatives (*thing, one, time, place, way, reason*)
- (v) *all*

All clefts and pseudo-clefts are identifying constructions which, with the exception of system-deviant examples, contain a form of the copula. All clefts, as defined in this book – see Section 3.2 – contain predicative *it*. All pseudo-clefts, as defined in this book – see Section 3.1 – contain a relative clause, either introduced by one of the *wh*-words that function as interrogatives in English, or headed by one of their pro-form equivalents, or by *all*.

3.1 THE CLASS OF PSEUDO-CLEFTS

Notes

- (ii) 3 and 5 tend to be mutually exclusive.
 (iii) The structural change required to convert a basic pseudo-cleft, so represented, into the corresponding reversed pseudo-cleft, would be: 1 2 3 4 5 6 \Rightarrow 6 3 4 1 2 (5 is not represented in reversed pseudo-clefts, as we see from the ungrammaticality of *'Not only that was why he left').

Table 3.1 Types of relative clause in pseudo-clefts in LL + LOB

	Fund		Lexically headed							Total	
	Basic	Revised	Total	The	Adj	Head	Rel	Basic	Revised	Total	
what	187	302	789 (78.4%)	+	-	thing(s) thing(s) thing(s) thing(s)	- + + -	28 11 28 12	98 6 14 1	687 (34.8%) (66.5%)	
who				+	-	one(s) one(s) one(s)	- + +	2 6 1	11 15 3	42 (14.9%) (4.1%)	
why	1	90	91 (12.1%)	+	-	reason reason reason	- + +	6 7 2	2 3 3	19 (9.9%) (11.5%)	
how		26	26 (3.5%)	+	+	way way way	+	2 1 5	11 1 3	25 (8.9%) (4.9%)	
where		39	39 (5.2%)	+	-	place place place place	- + - +		1 1 1	4 (1.4%) (4.2%)	
when		6	6 (0.8%)	+	-	time time time	- + +		1 1 2	5 (1.8%) (1.1%)	
Total	293 (38.3%)	463 (61.7%)	751 (100%)	+	+	time adj obj	+	45 14 176	16 5 106	80 (28.4%) (100%)	

A contrived example which represents perhaps the most extended structure that we could expect to find is as follows:

- (1) The thing for which he was famous could not have been simply his sense of the absurd.

3.1.1. Wh-clefts and th-clefts²

Some may question the parallels posited above between *wh*-clefts and *th*-clefts on the grounds of the admittedly dubious acceptability of many basic *wh*-clefts with a *wh*-item other than *what*. However, the parallels are persuasive if we consider reversed pseudo-clefts with a demonstrative as subject/themc. Compare:

- (2) a. ?*Who took your purse was Mary.
b. The one who took your purse was Mary.
c. That's who took your purse.
d. That's the one who took your purse.
- (3) a. ?How I did it was by rotating the knob.
b. The way I did it was by rotating the knob.
c. That's how I did it.
d. That's the way I did it.

Speakers may dispute the acceptability of the sentences in a., but none would surely reject the reversed *wh*-clefts in c.

If we allow the class of pseudo-clefts to include members whose relative clauses are headed by forms such as *one* and *place*, should we allow those with nouns such as *man*, as in the following corpus example?

- (4) # Frank Morgan was the man who . STARTED all this # (LL
S.2.4,184)⁴

Before answering this question let us draw a distinction between two classes which, following Halliday and Hasan (1976), we may call 'pronouns' (*thing, one, place, way*, and so on) and 'general nouns' (*thing, person, man, place, kind*, and so on), with *thing* and *place* belonging to both classes. A major difference between pro-nouns and general nouns is that the latter, but not the former, may function cohesively like anaphoric reference items. Compare the following:

- (5) a. It is unwise to leave your car unlocked. The thing could easily be stolen.

- b. The incidence of muggings in the neighbourhood has declined.
 c. The thing that disturbs us is the rate of car thefts.
 (6) a. Have you been to Adelaide? I love the place.
 b. I left Sydney last Wednesday. The place I am heading for is Adelaide.

Thing and *place* in the a. sentences function like anaphoric reference items (which explains why (5a) and (6a) would sound odd if *thing* and *place* were assigned nuclear stress). In both cases it would be possible to substitute *it* for 'the + general noun'. By contrast, in the b. sentences, pro-noun *thing* and *place* are not anaphoric: there is no presupposition of an earlier occurrence of a word such as *object*, or *location*.

Halliday (1967a:234 and, with Hasan, 1976:103) claims that pro-noun *thing* is restricted to non-human and indefinite nouns (thereby contrasting with *one* which, as a pro-noun, is restricted to definite human nouns). *Thing* can, however, have a definite human referent, as the following corpus example shows:

- (7) # - [a:m] the only thing they have in common is the children they've PRODUCED # (LL S.5.10,328)

Both *one* and *thing*, as pro-forms, have a plural as well as singular form:

- (8) # [ə] the ones I know BEST # are the poems of The Waste Land PERIOD # [LL S.3.1, 1326]
 (9) # those are the things that Miranda doesn't LIKE # (LL S.4.3,588)

Returning to our question regarding the example in (4), we might consider whether the presence of the general noun *man* contributes such an amount of lexical content that the semantic equivalence between the pseudo-cleft and the simple, or 'non-cleft', sentence which corresponds to it is impaired. In other words, is the relationship of (4) to the non-cleft *Frank Morgan started all this* the same as that of *Frank Morgan was (the one) who started all this* to the non-cleft? For most speakers the answer is probably 'no': true equivalence is possible only if the headword is a lexically empty pro-form.

A further difficulty that would arise should we accept as *th*-clefts those constructions with relative clauses headed by general nouns (though it is a practical rather than a theoretical difficulty), is the

open-endedness of the class. As Halliday (1967a:234), who allows general nouns in the context of pseudo-clefts, admits:

it is here that it is difficult to delimit the range of identifying clauses; perhaps any noun that dominates the identifier in the lexical taxonomy should be admitted.

Once a decision has been made to exclude general nouns, examples such as (10) and (11) are automatically excluded by virtue of the specific lexical content of *artist* (by comparison with *one*) and *car* (by comparison with *thing*).

- (10) The artist he resembles most closely in spirit is, I think, Watteau. (LOB CO7,67-8)
 (11) # - - and that was the CAR # - which I [ə sto ə] which I allowed to go BY # (LL S.12.4,502-3)

One potential objection to the inclusion of *th*-clefts in the class of pseudo-clefts is that they may feature (theoretically) unlimited modification of the head. However, when the head is, for example, modified by a single numeral or quantifier, as in (12), (13) and (14),

- (12) # that was the first thing they DID # (LL S.1.11,610)
 (13) # and (at) the only time I can RELAX is when I ACCEPT this # (LL S.7.1,1570)
 (14) The second thing which the university does is to give its students a special experience in which they gain an abiding insight into a university's perspective. (LOB G58,81-3)

a non-cleft version is readily available (*they did that first: I can relax only when I accept this; secondly, the university gives . . .*). With other classes of modifier (such as evaluative epithets), and with complex modification, it would be more difficult to make out a case for inclusion in the class of pseudo-clefts, since there is no non-cleft version available. For this reason, examples such as the following have not been included in the corpus:

- (15) The best thing I can do is lie still. (LOB N23,86)
 (16) You, Juarez, are about the only one whose loyalty I can take for granted. (LOB N19,39)
 (17) Said Cavanagh, 46 and for years the favourite designer of Princess Alexandra and the Duchess of Kent: 'This is one of the most thrilling things that has happened to me since my shop opened nine years ago.' (LOB A09,231)

3.1.2 *All-clefts*

Most people would agree that there is a close semantic relationship between the following examples – which I am claiming to represent a subclass of pseudo-clefts, *all-clefts* – and the *th-clefts* with relative clause headed by *the only thing* which correspond to them.

(18) # then all we need is a . TELEVISION # (LL S.4.1.581)

(19) # and all I'd do was get UPSET # (LL S.2.12.305)

(20) All I know is that I love him. (LOB K07.173)

The criterion by which *all-clefts* can be claimed to represent a subclass of pseudo-clefts is the set of proportionate semantic relationships which obtains between corresponding *wh-clefts*, *th-clefts*, and *all-clefts*, non-clefts. In the sentences below, (21a) is to (21b) as (21c) is to (21d):

- (21) a. What/The thing the car needs is a new battery.
 b. All the car needs is a new battery.
 c. The car needs a new battery.
 d. The car only needs a new battery.

The feature that is shared by (21b) and (21d), but absent in (21a) and (21c), is an assertion of exclusiveness associated with the presence of *only* (actual in (21d), implied in (21b)).

There is, in addition, a feature shared by the pseudo-clefts in (21a) and (21b), but absent in (21c) and (21d), their non-cleft counterparts. Pseudo-cleft (and cleft) constructions have often been noted to carry an implicature of exclusiveness, or exhaustiveness (see further Section 4.3 below). Thus if we compare (21a) and (21c), for example, we note that whereas (21c) envisages the possibility that the car needs other things as well as a battery, (21a) equates the car's need exhaustively with a battery. (21d might, in other words, be regarded as something like an implicature licensed by (21a).)

Unfortunately, the types of sentences one may construct to test the exclusiveness implicature elicit variable responses when submitted to the judgment of native speakers. The acceptability symbols applied to the following sentences are therefore offered tentatively.

- (22) a. The car needs a new battery, {amongst other things.
 and it needs a new
 alternator too. }

- b. *The car only needs a new battery, {amongst other things.
 and it needs a new
 alternator too. }
- c. ?It is a new battery that the car needs, {amongst other things.
 and it needs a new
 alternator too. }
- d. ? {What
 The thing } the car needs is a new battery, {amongst other things.
 and it needs a new
 alternator too. }
- e. * {All
 The only thing } the car needs is a new battery, {amongst other things.
 and it needs a new
 alternator too. }
- f. *It is only a new battery that the car needs, {amongst other things.
 and it needs a new
 alternator too. }

The unacceptability of (22b), (22e) and (22f) seems to be associated with the presence of the *only*: exclusiveness is part of the proposition being asserted. Of the remainder, (22a) is universally accepted, but (22c) and (22d) are rejected by many. This rejection can only be attributable to the exclusiveness implicature associated with cleft and pseudo-cleft constructions.

The two kinds of exclusiveness may be differentiated as follows. The 'asserted exclusiveness' which is conveyed by *only* is part of the propositional semantics (or 'ideational', in Hallidayan terminology). As such, it is subject to 'universal' questions of truth and falsity, and affected by such grammatical processes as negation and interrogation. Thus, for example, (23a) asserts something quite different from (23b). In the latter the exclusiveness can be contradicted:

- (23) a. The car only needs a new battery.

- b. The car doesn't only need a new battery.

(The 'implied exclusiveness' conveyed by clefts and pseudo-clefts, as identifying constructions, is a matter of pragmatics: an inference warranted by the construction. Because it is not part of the propositional semantics, this feature is not affected by negation or interrogation. Thus the implied exclusiveness of (22c) is retained in its negated counterpart, (24):

- (24) It is not a new battery that the car needs.

3.2 CLEFTS

(In considering the syntactic structure of clefts it is helpful, as in the case of pseudo-clefts, to begin with the formula presented by Prince (1978:883):

It is/was C_i which/who(m) that Ø S - C_i

There are several reasons why it is necessary to modify this formula.)

- (a) The superordinate clause may select for modality, aspect and polarity, and may include a 'focusing adverb' (*only*, *just*, and so on) between the copula and the highlighted element, as the following examples attest:

- (25) It may have been then that Trelawny contrived to do his copying. (LOB G07,38)
 (26) it is only on a minority of issues that the House of Commons can formulate an independent view. (LOB G53,138-9)
 (27) # [ə] it's not that Mervyn's [ə² ə²] TOTALLY unreliable # (LL S.2.6.1119)

- (b) The class of relative *wh*-words must be expanded to include *when* and *where*, and also *wh*-words serving as complement to a preposition, as in *in which* and *for whom*.

Figures for the different items in the corpus are given in Table 3.2, which shows that prototypical clefts with relative clause introduced by *that* account for about two-thirds of all tokens. There are no tokens with *where* in the corpus.

- (c) The absence of a highlighted element with experiential function in cleft sentences such as (27) above, and (28) and (29) below,

Table 3.2 Types of relative item clefts in LL + LOB

	Number	%
<i>that</i> ¹	483	64.2%
<i>which</i>	52	6.9%
<i>who</i>	91	12.1%
<i>when</i>	14	1.9%
Ø	108	14.4%
Prep + <i>which</i>	4	0.5%
Total	752	100%

¹ The figures for *that* include one example of *wor* ((49) in this chapter).

requires that the highlighted element be characterized as optional.

- (28) # [i] is it that you you're not in a different PÖST THERE # (LL S 2.7,410)
 (29) # - it may well have BEEN # that we'd have been able to help HER to SEE # - the damaged bit of HER # that . SOMEHOW # COLLUDED # with the damaged hit of YÖU # (LL S.6.5,527-32)

That these examples are in fact clefts rather than structures deriving from extraposition of the nominal clause requires some justification. Unfortunately, convincing evidence is difficult to find. One piece of evidence is the typical availability of a non-cleft counterpart for such sentences ('Mervyn's not totally unreliable', 'Are you not in a different post there?', 'We may have been able to help her to see . . .') but not of a non-extraposed counterpart (*That Mervyn's totally unreliable is not', *Is that you're not in a different post there?', *That we'd have been able to help her to see may well have been'). Of course not every such example does have a non-cleft counterpart. For example the non-cleft version of (30)

- (30) it may be that a frontal view will be more effective in certain circumstances (LOB E10,92-3)

is the ungrammatical *A frontal view may will be more effective in certain circumstances'. However, if the non-cleavability of such cases constitutes a problem for the cleft interpretation, then it is no less of a

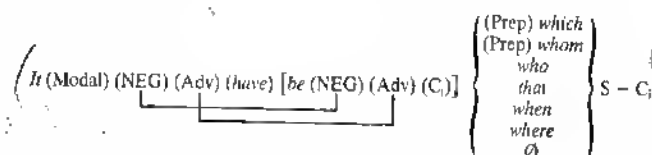
problem for an uncontroversial cleft such as (31), whose non-cleft version is the equally ungrammatical *Very seldom will can good be done in this field under six months':⁵

- (31) It will be very seldom that permanent good can be done in this field under six months (LOB H08,50-1)

The thematic evidence in favour of a cleft interpretation is, like that relating to non-cleft counterparts, suggestive rather than decisive. The thematic prominence that the sentences under consideration assign to the non-ideational item(s) following the *it* is arguably more compatible with a cleft interpretation than an extraposition explanation. Whereas the communicative *raison d'être* of clefts is unquestionably thematic/informational (see Sections 5.2.3 and 5.3.5 below), that of extraposition structures is generally interpreted in terms of the principle of 'end-weight' (see Section 1.3 above).

Delahunty (1984:71) analyses the *that*-clause of sentences such as (25) and (27) as the highlighted element rather than the relative clause. Such an interpretation is inconsistent with their discourse behaviour (which is neither like that of elliptical clefts – see further Section 3.4 below, nor like that of simple copular sentences with referential *it* as subject (see further next Section)). It follows that, if sentences such as (25), (26) and (27) are accepted as clefts, one of the nine 'properties' of cleft sentences listed by Delahunty (1984), and which is assumed in most analyses (namely 'The clause always contains a gap, or trace, with which the focus is (ultimately) associated', p.65) is not valid.

In light of these observations, we may recast Prince's formula as follows:



Notes

- (i) Only one occurrence of NEG and Adv is possible
- (ii) If the items in square brackets are absent, only *that* or \emptyset may be selected.

3.3 IDENTIFICATION VERSUS ATTRIBUTION

In Section 1.2 above it was claimed that clefts and pseudo-clefts need to be distinguished from constructions which have a similar sequence of elements but which are to be interpreted differently, as attributive constructions. Because the distinction between identification and attribution is fundamental to the task of defining the class of pseudo-clefts, and is arguably also relevant to defining the class of clefts, I shall begin this section with an examination of a set of non-cleft examples which will facilitate the task of establishing the distinction. Compare the sentences in (32) and (33):

- (32) a. The youths on bikes are a vicious gang.
b. A vicious gang are the youths on bikes.
(33) a. The youths on bikes are the local gang.
b. The local gang is the youths on bikes.

The sentences in (32) are attributive, with a property being ascribed to the subject, *the youths on bikes*, in both cases. The difference between them is merely thematic: (32b) displays fronting of the complement, *a vicious gang*, and displacement of the subject to post-verbal position. The sentences in (33) are identifying, with *the local gang* (assuming that *gang* is stressed in both cases) serving to define the identity of *the youths on bikes*. Following Halliday (1967a, 1985) I shall refer to *the local gang* as the 'identifier' and *the youths on bikes* as the 'identified'. By contrast with the sentences in (32), (33a) and (33b) display more than merely an ordering difference. Here the grammatical functions differ as well: *the youths on bikes* is subject in (33a), but complement in (33b). The syntactic difference between the sentences in (32) and (33) is reflected in the fact that there is only one *yes/no* interrogative corresponding to (32), namely 'Are the youths on bikes a vicious gang?', but two for (33), namely (for (33a)) 'Are the youths on bikes the local gang?' and (for (33b)) 'Is the local gang the youths on bikes?' Furthermore, the syntactic difference between (33a) and (33b) is reflected in the different forms of the copula required by the plural and singular subjects respectively.

Identification then, as noted in Section 1.2, is a relationship between two entities, whereas attribution is a relationship between an entity and some attribute that is ascribed to it. The fundamental grammatical difference between the two modes is reversibility: identifying clauses are reversible, as the sentences in (32) and (33) indicate, but attributive sentences are not (what appears to be a case of reversal with the attributive sentences in (32) turning out to be merely a case of thematic ordering).

It is the property of reversibility which, above all, enables us readily to distinguish pseudo-clefts from superficially similar attributive constructions. The latter include not only sentences of the type in

(2) in Chapter 1, reproduced for convenience below as (34a), but also those with an embedded interrogative clause, both the *wh*-type as in (34b) and the *yes/no*-type as in (34c):

- (34) a. What Tom offered Sue was too sweet.
b. What Tom offered Sue is unknown.
c. Whether Tom offered Sue a sherry is unknown.

(Unlike genuine pseudo-clefts, the sentences in (34) cannot be reversed. Nor can any of the sentences in (34) be unclaved (for example *'Tom offered Sue too sweet').)

Although no examples of sentences ambiguous between a pseudo-cleft (identifying) reading and an attributive reading were encountered in the corpus, it is not difficult to invent examples. Consider (35):

- (35) What Tom offered Sue was a puzzle.

(35) is ambiguous between an identifying reading (on which it qualifies as a pseudo-cleft), where a *puzzle* is understood in a substantive, referential sense (for example, 'a crossword puzzle'), and an attributive reading, where the embedded *wh*-clause is interrogative and a *puzzle* has adjectival force. On its identifying reading (35) can be reversed, and shares a close semantic relationship with the corresponding non-cleft 'Tom offered Sue a puzzle', cleft 'It was a puzzle that Tom offered Sue,' and right-dislocated agnate 'It was a puzzle, what Tom offered Sue.' On its attributive reading (35) cannot be reversed or unclaved, and has no cleft or right-dislocated counterpart. On the other hand it does, unlike the sentence on its identifying reading, have an extraposed version ('It was a puzzle what Tom offered Sue'). It is because (35) does not, on its attributive reading, in any sense represent a 'claving' of a simpler clause that it can be excluded from the class of pseudo-clefts.

Consider another example:

- (36) What Tom is is ridiculous.

(36) has both an identifying reading (on which it qualifies as a pseudo-cleft) and an attributive reading. On the attributive reading the sentence does not identify the referent of the *wh*-clause, but merely says something more about it (with *What Tom is* referring, for instance, to Tom's occupation, or to his role in a theatrical performance). On this reading the sentence has no cleft or non-cleft counterpart, but there is an extraposed version in 'It is ridiculous what Tom is'. On the identifying reading the sentence is closely related semantically to 'Tom is ridiculous' and 'It is ridiculous that Tom is'.⁶ The

sentence is identifying, even though both the *wh*-clause and the highlighted element are attributive, precisely because the structure as a whole is identifying, specifying an identity for an entity that is to be identified.

The constructions which are superficially similar to pseudo-clefts that I have examined so far have involved intensive attribution. Occasional examples may also be found of sentences, superficially similar to pseudo-clefts, which involve circumstantial attribution. Consider a sentence that is ambiguous between a (circumstantial) identifying reading (on which it qualifies as a pseudo-cleft), and a (circumstantial) attributive reading (on which it is not a pseudo-cleft):

- (37) The place where Tom made the offer was in Perth.

On the identifying reading the place that is referred to is understood to be Perth: two circumstantial elements of location are equated. On the attributive reading the place is understood as being in Perth: a circumstantial attribute (that of 'being located in Perth') is ascribed to an entity (*The place where Tom made the offer*). The two readings are distinguishable via the tests of unclavability (only on the identifying reading is (37) unclavable to 'Tom made the offer in Perth') and reversibility ('In Perth was the place where Tom made the offer' is a true reversed form of the identifying reading, requiring a different probe from the non-reversed form, but merely a thematic variant of the attributive reading, with the *The place where Tom made the offer* still serving as subject).

As with pseudo-clefts, so with clefts, there are constructionally similar attributive sentences. Here, however, the case for excluding them from the class of clefts is less clearcut and compelling. In fact, I shall argue, while some may be discarded, there are good reasons for accepting others as clefts.

One type of attributive construction that is to be rejected from the class of clefts is exemplified by (3) in Chapter 1 above, and the italicized sentence in (38).

- (38) >B (. . .) # --- I've HEARD of Casa Pupa # why have I
heard of Casa PUPA #
A it's a lovely shop that has SPANISH things #
(LL S.2.11.338)

The subject *it* in (38) functions as a genuine anaphoric pronoun (its antecedent being *Casa Pupa*), and the *that*-clause serves as a dependent relative within the noun phrase headed by *shop*. It can be seen that this sentence is not a cleft by substituting a non-cleft sentence, 'A lovely shop has Spanish things'. This does not mean the same thing as the original sentence, and is inappropriate in the context.

40 Cleft and pseudo-cleft constructions in English

The second type of attributive construction that I am here rejecting from the class of clefts is the proverbial type discussed by Jespersen (1937:89).

- (39) a. It is a poor heart that never rejoices.
b. It is an ill bird that fouls its own nest.
c. It is a long lane that has no turning.

The meaning of (39a) Jespersen takes to be 'The heart that never rejoices is poor' rather than 'That which never rejoices is a poor heart'. In other words (39a) is a generic statement ascribing the attribute ('being poor') to any 'heart that never rejoices'. A more semantically transparent form of (39a), with 'poor' alone as highlighted element, would be:

- (40) ?It is poor that the heart that never rejoices is.

The unacceptability of (40) is reduced marginally if a contradicting expression is added ('It is poor that the heart that never rejoices is, not indifferent'). Sentence (41), from LOB, belongs in the same category as those in (39). It has a proverbial flavour, and is clearly attributive rather than identifying, being equivalent to 'A man who believes that the contrast . . . is (a) foolhardy (man)'.

- (41) It is a foolhardy man, surely, who believes that the contrast had nothing to do with the expression of the tradition through, not only the Monarchy as an institution, but also the personal characters and examples set by George V and George VI. (LOB G59,99-102)

Declerck (1983b) argues convincingly that 'proverbial' sentences of the type in (39) and (41) are not in fact clefts, but represent 'a type of sentence that is homophonous with clefts' (p.14). The main piece of evidence that Declerck points to is that the *wh-/that*-clause of sentences like (39) and (41) is a genuine restrictive relative clause, whereas the *wh-/that*-clause of clefts differs from a true relative clause in several respects. For instance, like restrictive relative clauses with non-specific noun phrase antecedents, but unlike the relative clause of clefts, the *wh-/that*-clause of 'proverbial' sentences like (39) and (41) yields a conditional interpretation. Sentences (39a) and (41) are thus interpretable as (42a) and (42b) respectively:

- (42) a. If a heart never rejoices, it is (a) poor (heart).
b. If a man believes that the contrast had nothing to do with the expression of the tradition . . . , he would be (a) foolhardy (man).

Defining the class 41

The *it* of 'proverbial' sentences is arguably to be analysed as a determinative pronoun.⁷ One piece of supporting evidence is the possibility of replacing *it* by *he* or *she* where appropriate (for example 'He is a foolhardy man, surely, who . . .'). A possible objection, that determinative *it* cannot normally have a human referent, disappears if it is kept in mind that the type of sentence under consideration is, as Declerck (1983b:15) notes, 'essentially proverbial and, like most proverbs, reminiscent of an older stage in the English language'. In earlier English, as the following examples indicate, *it* could be used as a determinative pronoun with postponed relative clause and could be used in attributive sentences where today *he/she* would be preferred:⁸

- (43) a. It holds current that I told you of.
b. 'Tis a good boy', said his master.

There are constructions which, though they display some attributive features, may be considered clefts. Examples are (44) and (45):

- (44) well it's . mainly a LARKISH course # - that [a] we're MAKING # (LL S.9.2.638-9)
(45) It is not a sentimental, but a precise point which he makes: (LOB G59,93-4)

I shall regard these examples as clefts, despite the fact that they seem to be close in meaning to sentences that are clearly attributive ('The course that we're making is mainly Larkish', 'The point which he makes is not sentimental but precise'). It is not the whole highlighted element that is attributive in either case, but merely *Larkish* in (44), and *sentimental* and *precise* in (45). The reason why the sentence is felt to be attributive in each case is that only the attributive element represents new information: all the rest is backgrounded. If the noun head (*course*, *point*) were not presupposed, it would be stressed and the sentence would be unambiguously identifying. It is not in fact possible to interpret the whole of the highlighted element as attributive, as can be seen from the ungrammaticality of (46), where the entire highlighted element serves as subject complement in the relative clause.

- (46) *it's mainly a Larkish course that the course that we're making.

Declerck (1983b) claims that sentences of the type presently under discussion, which he labels 'predicational clefts', are 'really borderline cases: although the information they convey is predicational ["attributive", P.C.], they are formally cleft sentences, and cleft sentences are essentially specificational ["identifying", P.C.] in nature' (p.18). One characteristic that clefts like (44) and (45) share

with attributive sentences in general is that they admit comparison and modification of degree (whereas identifying sentences do not):

- (47) It is a rather more precise point which he makes.

Further evidence is the possibility of substituting *no* for *not* *a*, permissible only if the negated noun phrase is attributive and not identifying. Compare the acceptability of (48a) and (48b):

- (48) a. It is no sentimental point which he makes.
b. *It is no surgical appliances which he makes.

On the other hand there are respects in which sentences like (44) and (45) behave like identifying structures. They have a noun phrase as identified element (albeit a presupposed noun phrase: *course* in (44), and *point* in (45)), and this element has a specific reference which makes it compatible with the implicature of exclusiveness which is associated with the cleft construction (see Section 4.3 below).

In addition to clefts in which the highlighted element and relative clause are related through intensive attribution, as in (44) and (45), there are those where the relationship is possessive and circumstantial attribution. The examples in (49) and (50) have a focal possessive item modifying the head noun.

- (49) Mrs X, the carpenter's wife, died. There were two Mrs X's in the village. Rumour at first reported the wrong one, at which Mr X, the carpenter, was deeply incensed.

'It's my wife who's died. Surely I ought to know,' he said.
(LOB G10,157)

- (50) (...) # I'm NOT appealing to our friends HERE # - THEY want to help # in {EVERY} possible WAY #, but I UNDERSTAND # that we've been HONOURED # by the visit of several DISTINGUISHED #, pro-Market EDITORS #, for this DEBATE #, with those EDITORS HERE #, it is THEIR credibility that's in QUESTION # (LL S.12.5,429)

The head of the highlighted noun phrase is presupposed in each case, yielding the following paraphrases: 'The wife who's died is mine'; 'The credibility that's in question is theirs'. Copular sentences of this type, with possessive noun phrase as complement are, as Halliday (1985:121) has observed, inherently ambiguous between attribution ('the wife who has died is a member of a class of my possessions' and so on) and identification ('the wife who has died is identified as belonging to me'). Taglicht (1984:57) has observed that clefts of this type have, as stylistic alternatives, clefts with the nominative or

accusative case form of the possessive pronoun (or case unmarked form of the possessive noun) as highlighted element, and *whose* as relative pronoun:

- (51) a. It is me whose wife's died.
b. It is they/them whose credibility is in question.

Taglicht (1984:57) makes the claim, which is supported by the findings of the present study, that the type of cleft represented in (51) is less likely to occur than that represented by (49) and (50), in 'unaffected speech'.⁹ The former, on the other hand, are more likely to be found in formal writing, where the coextensiveness of focal item and highlighted element may be an advantage. (The pragmatic import of the normal printed form of (50) could be clarified only by the context, unless the device of underlining or italics were used.) Although the two patterns are sufficiently close in meaning for the choice between them to be regarded as largely stylistic, they are not totally equivalent. In (50) *their credibility* represents the 'identifier', and the 'identified' constituent, *that is in question*, carries the presupposition 'something is in question'. In (51b) *they* represents the 'identifier', and the 'identified' constituent, *whose credibility is in question*, carries the presupposition 'someone's credibility is in question'. In (50) the focus of information *their* represents but a subsection of the identifier, but in (51b) *they* represents both the focus of information and identifier.

The highlighted element and relative clause of the clefts in (52) and (53) are related through circumstantial attribution.¹⁰

- (52) For Joyce as much as for Proust, it was the 'I', the moi, with which he was ultimately concerned. (LOB G41,100-1)

- (53) It is with this aspect that we should be mainly concerned. (LOB B15,146-7)

The semantic relationship between highlighted element and relative clause in (52) and (53) seems best regarded as between an entity ('his ultimate concern', 'our main concern') and a circumstantial attribute of matter that is ascribed to it ('with the "I", the moi', 'with this aspect'). There is not an equation between two like participants that is the mark of an identifying structure. Nevertheless, as is customary in copular sentences of the circumstantial type, the dividing line between the attributive and identifying modes is not absolutely clear-cut. It is not obvious whether the *with*-phrases designate a class of things (whose members are all concerned with the 'I', the moi, or with this aspect), or an identity (the thing that is identified by 'being concerned with the "I", the moi', or 'being concerned with this aspect'). The difference between (52) and (53) is similar to that

between (49) and (50), and (51a, b) respectively. This is that the type represented by (52) is, like that represented by (51), by and large restricted to writing.¹¹ In this case, however, the likely explanation is not the absence of accent marking in writing, but the sensitivity of many writers to the prescriptive rule forbidding sentence-final prepositions.

Predicational clefts with highlighted adjective phrase are favoured in Celtic-based varieties of English.¹² There are two examples in LOB from Scottish English, (17g) in Chapter 4 below and (54):

- (54) 'I'll have to ask him. What do the people here think of Martiu?'
'Och, it's *hide enough* they've seen of him. He's not been here more than a few days altogether. They don't mind him. he's a harmless enough wee man.' (LOB L11,144)

3.4 SYSTEM-DEVIANCE AND INCOMPLETENESS

There are methodological problems that the 'total accountability' principle of the corpus-based approach makes it necessary to consider. Not only is it important to delimit precisely the class of 'system' constructions under attention, but we need to take account of some of the ways in which factors such as language planning and processing, speaker-interactive variables and other situational and discourse phenomena can result in the encoding of system-deviant, or at least untypical, constructions.

The system-deviance of several pseudo-cleft tokens in the corpus results from omission of the copula. For instance in (55) the dependent status of the clause which realizes the last tone unit indicates that it constitutes the highlighted element of a pseudo-cleft lacking the copula, rather than a semantic extension of an incomplete pseudo-cleft consisting of *what*-clause only.

- (55) # but - it must have been some {HÜGE} AFFAIR #
because - *what* {FÄN didn't} REALIZE # *not being a*
LAWYER # . or a lawyer's WIFE # - *that apparently the WIFE*
was put on probation # (LL S.1.13,685-8)

Sometimes an anacoluthon will result in a cleft or pseudo-cleft that contains two copulas. The speakers in (56) and (57) repair the tense of the copula from present to past.

- (56) # it's really was BERYL that did it I THINK # (LL S.1.5,401)
(57) # . [ə ʔ ʔ ʔ] that's in FACT # - [ə] WAS # [əm əh] what she was . putting FORWARD # at that TIME # (LL S.12.3,1004-7)

Processing factors may be responsible for system-deviance. For example in (58) the infinitive *to know*, though grammatically in construction with *'m fascinated*, is encoded as part of the highlighted section of the pseudo-cleft.

- (58) # - now what I'm FÄSCINATED # is to know that
COMMUNISTS # (coughs) and FÄSCISTS # are such bad
SHOTS # . that WIJEN they have # . a battle between
THEMSELVES # . ALL their MISSILES # hit the
UNFORTUNATE {Mr BRÖÖKE #} # who as I
UNDERSTAND # is NEITHER # (LL S.5.4,178-87)

There are, broadly speaking, two categories of incomplete pseudo-clefts and clefts: those whose completion is prevented by a variety of contextual or processing factors, and elliptical constructions in which omitted material is textually recoverable. I shall discuss such cases respectively under the headings of 'interruption' and 'ellipsis'.

(i) Interruption

An occurrence in the situation may disrupt the speaker, as in (59), where Speaker A is distracted by a problem that he encounters when he twists a microphone whose operation he is exploring.

- (59) A good HĒAVENS # - . *what happens when you TWIST it* # is
that oh JĒSUS #
a oh nothing it's all right (LL S.1.7,702-3)

An interlocutor may interrupt the speaker, with the result that a cleft or pseudo-cleft is left incomplete:

- (60) B YĒAH # [ə:m] well this was {ðə ʔ əs} a scheme which we put
forward to PĒRRINS # just about a year and a half AGŌ #
-[ə:] . to Sir Alec WADDĒLL # .
a that's it . that is Perrins then
B [ə:M] . [ə:]
a that is it . it was he who/s/
B YĒS # YĒS # [mh̃n] # (LL S.2.1,210)

A number of cases in this category result not from an interruption in the literal sense of the term, but from an 'interruption' that originates with the speaker, whose failure to complete a construction that has been embarked upon results from becoming sidetracked. An example is (61):

- (61) a well we had some pretty dreary old men in the First World
War didn't we
A the the best [ə:] certainly *what Britain did in the SĒCOND*
World War was [ə:] # I mean I don't KNŌW # I merely

REPEAT # what I've read in BOOKS # so I don't really
KNOW # kind of

B YES # -

A can't TESTIFY that # . but in one very good [te] the British
army was very much BETTER in some respects # in the
{SECOND World War} than in the FIRST # (LL
S.2.3,231)

A further type of factor that may result in an incomplete construction being recorded in the transcription is speaker incomprehensibility or inaudibility, as in (62).

(62) A [m] # . YEAH # that's the NICE thing # about having a
SEASON ticket

B *that's what's good about having* (murmurs) (LL S.7.2,807)

(iii) Ellipsis

In some linguistic contexts an elliptical interpretation is clearly called for. For instance in (63) the relative clause of the italicized cleft is ellipsed under conditions of direct recoverability from the relative clause of the immediately preceding cleft.

(63) It is not the observation of likenesses which is at fault in popular etymology, *it is the fact that conclusions about the relationships of words, drawn from comparisons, happen to be erroneous.* (LOB G51,59-61)

More problematical is an example of the type in (64), which raises the question: do we have two clefts here, the second with ellipsis of *it* and the copula, and of the relative clause; or, do we have only one cleft, with paratactically related highlighted material displaced to the end of the relative clause? It is the elliptical interpretation that has been adopted in the present study, on the grounds of the strong resemblance between such an example and that in (63).

(64) It's their interest you want - not their sympathy. (LOB F03,175,176)

The criterion that has been used to separate clefts with ellipsed relative clause, such as (63), from copular sentences with referential *it* as subject, is the clear retrievability of items representing the ellipsed relative clause, either directly from the co-text or context, or indirectly via inferences from them. To put it another way, a non-elliptical interpretation has been preferred unless there is strong evidence indicating ellipsis.

The simple copular sentences which have not been accepted as elliptical clefts in the present study are of two types. In the first, *it*

refers anaphorically or deictically to a sense perception, as in (65), or to a general notion (referred to by such 'operator' nouns as 'problem', 'trouble', 'thing', 'reason', 'cause', and 'question'), as in (66) below.¹³

(65) 'The cat will have got itself out through the coal-shoot. Bound to -'

'It hasn't. I heard it mewling. I am sure and certain *it was the cat* - let go of me, George!' (LOB P01,105-6)

If the italicized clause in (65) is elliptical, then its complete form would be 'it was the cat mewling' (with *it* referential to an implied noun phrase, 'the noise'), rather than the cleft 'it was the cat which was mewling'.

Declerck (1981) argues convincingly that sentences such as 'it was the cat mewling' and 'it was someone screaming for help' are not clefts. (Unlike clefts they do not, for example, alternate with pseudo-clefts, cannot use *that* after a proper name, and do not allow deletion of the relative pronoun.)

An example of a cleft-like sentence with *it* anaphoric to an implied 'operator' noun is (66):

(66) it's going so SLOWLY though you know # it's this it's these
awful these awful SYMBOLS # (LL S.2.1,18)

The *it* of the italicized construction in (66) appears to be anaphoric to an implied noun phrase such as 'the problem' or 'the reason'. As in the case of (65), if the construction in question here is considered elliptical, then the full form would be something like 'it's these awful symbols which take so long to type'. In this, as in the original sentence, it is a referential pronoun, rather than the predicative *it* of the cleft construction.

The second type of copular sentence that has been excluded from the present study is that with a predicate nominal having a human referent, as in (67).

(67) The word put an idea into his head, and he hurried through to the lounge and went to the phone. He dialled a Streatham number, and in a few moments was gratified to hear Conquest's clear voice.

'It's *me*, sir - Fred,' panted the porter. 'Something's happened, sir.' (LOB N05,178)

Constructions of this type have been much discussed in grammars of English. For instance Curme (1931:7) claims that *it* is here 'used as a subject to point to a person or thing that is first presented in only dim

outlines by the situation, but is often later identified by a predicate noun: "It's John" . . . (uttered by someone upon hearing approaching steps). That such constructions are reduced clefts is argued by Poutsma (1916), Jespersen (1958), and Declerck (1983a). None of these writers demands the ready retrievability of the relative that is required of copular sentences to qualify as elliptical clefts in this book. Poutsma (1916:732), for instance, states simply that the ellipsed clause should be 'more or less distinctly implied in the context'. Declerck (1983a) argues that all copular *it*-sentences that answer the question 'Who is it who . . . ?' are reduced clefts, and observes (p.241) that:

it is not necessary that the deleted relative clause should be recoverable from the *linguistic context* (i.e. should be identical with a phrase or clause that has been used before). In many cases it is the extralinguistic context or situation that makes clear how the reduced cleft should be completed. For example: (. . .) (103b) (talking about an engagement): Who is it [that she got engaged to]? (103c) (on seeing someone come in): Why, it's you [who are coming to see me]! What good news have you got to tell me?

While (67) would, then, be analysed as an elliptical cleft by linguists such as Poutsma, Jespersen and Declerck, the 'default' approach adopted here prevents the possibility of such an interpretation.

4 Formal properties

A number of analyses of cleft and pseudo-cleft constructions have been proposed to account for such aspects of their surface syntax as have been discussed in the previous chapter, as well as various 'connectedness' phenomena suggesting a close structural relationship between the highlighted element and relative clause. A detailed survey of these analyses would be beyond the scope of this book. In this chapter I shall indicate the type of analysis that is assumed in the study and examine several formal properties displayed by clefts and pseudo-clefts in the corpus (for which any adequate formal analysis of the constructions should be able to account). In subsequent chapters I shall offer explanations for a number of these properties in terms of the communicative functions of clefts and pseudo-clefts.

4.1 SYNTACTIC ANALYSES

For pseudo-clefts there is a substantial body of evidence supporting a direct syntactic description (as favoured, for example, by Quirk *et al.* 1985; Huddleston 1984; Higgins 1979; and Halvorsen 1978) in terms of independently established categories. In the case of transformational accounts this position is known as 'base derivation': the surface form of pseudo-clefts is regarded as similar to their underlying form, and it is therefore claimed that no specific clefting transformation is involved in their derivation. One piece of evidence, as noted by Green (1971), Morgan (1973) and Huddleston (1984), is that there are many cases of pseudo-clefts where there is no non-cleft counterpart that could provide a source for a cleft operation. An example follows (*that originally I was teaching by the direct method was stupid about it'):

- (1) # --- what was STUPID about it WAS that # ORIGINALLY
- [3] I was teaching by the direct METHOD#
(LL S.5.9,458-60)

Secondly, as noted in Chapter 3, the class of pseudo-clefts includes classes of constructions describable as *th*-clefts and *alt*-clefts which, in a formal syntactic analysis, it would not be possible to derive by cleaving. Even if it is maintained that *th*-clefts and *alt*-clefts are not pseudo-clefts, the similarities that they share, as identifying copular constructions, with *wh*-clefts should make us suspicious of any proposal to treat *wh*-clefts in terms of a radically different (cleaving) analysis. Huddleston makes the appealing suggestion that the different analyses required of cleft and pseudo-cleft constructions provide justification for the terminological distinction between 'cleft' and 'pseudo-cleft' (only the former being, in his view, satisfactorily described in terms of cleaving).¹

It is more difficult to settle on a satisfactory formal analysis of clefts. The types of considerations that necessitate the rejection of a cleaving derivation for pseudo-clefts do not apply in the case of clefts. Examples of clefts for which there is no corresponding non-cleft to act as a source are extremely rare. Furthermore there is the difficulty, in 'base derivation' accounts of clefts, that one is occasionally forced to postulate an ungrammatical source for either the superordinate or the subordinate clause. As Lees (1963) notes, in clefts of the type

- (2) a. It was of him that I asked it.
b. It was in the drawer that I put it.

there is no acceptable source for the superordinate clause *It was of him* in (2a), or for the subordinate clause *that I put it* in (2b).

While it is not clear precisely what steps are involved in the derivation of cleft sentences, some possibilities may be suggested. We may accept that the subject *it* of clefts is a non-referential dummy pronoun, but can we accept that its role is to serve as a trace for a subordinate clause that has been extraposed by contrast with, for example, the *it* of *It is possible that Tom offered Sue a sherry*? Analyses in which the subject *it* of clefts is taken as the by-product of a cleft-extraposition operation (with the relative clause thus serving as a postponed modifier to *it* in surface structure) are offered by, among others, Akmajian (1970), Emonds (1976), and Gundel (1977).

It is easier under this analysis than under those which construe the relative clause as modifying the highlighted element, to relate the cleft and pseudo-cleft constructions. Thus, under the postponed-modifier analysis, (3) below would be analysed as 'it - that stole her money while we were away in France - was Johnny' and related to 'The one that stole her money while we were away in France was

Johnny'. and (4) as 'it - that ventures out . . . - is so often the swimmer' (compare 'The one that ventures out . . . is so often the swimmer').

- (3) # *N* [əm] . it was JÖffNNY that stole her [əm] MÖNEY #
while we were A(WÄY) in FRÄNCE I THfNK # WÄSNT it
(LL S.2.13.1190-2)
(4) It is so often the swimmer that ventures out, gets into difficulties,
and is rescued, if there is time. (LOB B24,13-14)

Clefts with a referential human noun phrase as highlighted element and *that* as relative item (as in (3), (5) and (6)), which are quite common in the corpus, present difficulties for analyses in which the highlighted element is construed as antecedent. One problem is the interpretation of a defining relative as modifying an already-defined head, and another is the regular selection (see below) of *that* rather than *who* as relative item.

- (5) 'Kitty's all right', Bone contradicted flatly. 'It's her boy-friend
that's the trouble. If we could get rid of him . . .'
(LOB L04,37-8)
(6) 'If there's trouble it won't be me that starts it.' (LOB N14,170)

This postponed-modifier interpretation is open to several serious objections. One is that it fails to account for the agreement properties of the verb in the relative clause. The verb in the relative clause will always show number agreement with a highlighted nominal element, rather than with *it*, as illustrated in (7) below:

- (7) It is the tenants of the pre-1880 houses who are to be rehoused in
the new houses. (LOB J47,121-2)

It follows that it would not be possible to substitute *does* for *do* (or *salad* for *salads*) in (8):

- (8) # maybe it's SÄLADS {that DÖ it #} # (LL S.4.1.493)

There is some variation in person agreement, however, between pronominal highlighted elements and the verb in the relative clause. In (9) the verb in the relative clause, *starts*, does not agree in person with the highlighted first-person pronoun *me*.

- (9) If there's trouble, it won't be me that starts it. (LOB N14,170)

Another objection to the postponed-modifier analysis is that the subordinate clause of the cleft sentence is unlike an ordinary restrictive relative in a variety of ways. Although it resembles a restrictive relative clause in being introduced by the same set of pronouns that are used in relative clauses, the *wh*-forms are rare in comparison with *that* and zero (see Table 3.2 above). Not only are the comparative frequencies of the relative elements strikingly different from other relative constructions, but the range of syntactic functions that can be occupied by the relative element is greater. Another difference is that it is possible to delete a relative element functioning as subject in clefts, but generally not in other relative constructions.² Consider:

- (10) # [ɔm] well it's Bill Gravy wants to SPEAK to him # (LL S.8.1.833-4)

(Compare: *'Bill Gravy, wants to speak to him, is waiting outside'.)

- (11) # --- and [ə:] when [ə:] I think it was MRS Corley # TOLD your LORDSHIP # (LL S.12.3,1117-18)

Furthermore, unlike relative clauses, the subordinate clause of clefts may have as its antecedent not only a nominal element but an adjunct. Quirk *et al.* (1985:1387) comment that a construction without a noun phrase antecedent such as *It was because he was ill (that) we decided to return*,

makes inappropriate the use of the term 'pronoun' for the linking word *that*. It is noteworthy that a *wh*-relative pronoun cannot be used in cleft sentences where the focused element is an adjunct, and where consequently *that* does not have a strict 'pronominal' status:

**It was because he was ill which we decided to return.*

Amongst the suggestions for interpreting the surface structure of clefts that have been presented, that offered by Huddleston (1984:462) is one of the most challenging and, to date, least open to objection. Huddleston treats the *it* and *be* of the superordinate clause as grammaticalized features, and associates cleft sentences with their non-cleft counterparts by means of a cleaving operation. Huddleston regards the cleft construction as exhibiting subordination of the non-embedded kind (so that (1c) in Chapter 1, reproduced for convenience below as (12),

- (12) It was a sherry that Tom offered Sue.

would be taken to have as its immediate constituents *It was a sherry* and *that Tom offered Sue*).³ The antecedent of the relative clause is taken to be the highlighted element, but the relative clause and its antecedent are not considered to form a constituent.

This analysis avoids the difficulties which are encountered if we take the relative clause to be a restrictive modifier within a noun phrase structure headed by *it*. Huddleston's construal of the highlighted constituent, rather than *it*, as the antecedent for the relative element is supported by the 'connectedness' phenomena that have been noted in the literature. For instance, it has been observed by many that the distribution of reflexive pronouns functioning as highlighted element in clefts is the same as the distribution of reflexive pronouns in the corresponding non-clefts.⁴ Compare the sentences in (13) below (where *Tom* and *him/himself* are understood to be co-referential).

- (13) a. Tom wanted to pour a sherry for $\left\{ \begin{matrix} \text{himself} \\ *him \end{matrix} \right\}$.
b. It was for $\left\{ \begin{matrix} \text{himself} \\ *him \end{matrix} \right\}$ that Tom wanted to pour a sherry.

Furthermore, Huddleston's suggestion that the relative clause does not form a constituent with its antecedent reduces the difficulty of accounting for both clefts with a highlighted noun phrase, and those with other types of highlighted constituent, within a single analysis.

Huddleston's proposal is not without its difficulties. One is that it is, by his own admission, very largely *ad hoc*: the relative clause is of a type that is *sui generis*, unique to this construction. A further problem with the analysis is that if the subordinate clause of clefts is non-embedded, then it should be possible to tag it. However the acceptability of sentences such as the following is questionable:

- (14) a. ?It's the wife who decides, doesn't she?
b. ?It was Bill who had been approached, hadn't he?
c. ?It isn't unemployment that causes the most severe social problems, does(n't) it?

Nevertheless Huddleston's treatment of the cleft construction in terms of categories not all of which have been established on independent grounds, would seem to draw support from peculiar syntactic features of the construction (as described above, and explored further in Section 4.2 below).

In concluding this section it should be remarked that one of the major difficulties in arriving at an acceptable analysis of cleft constructions is the relative weight that different linguists will give to different kinds of evidence. For example the appeal of the postponed restrictive relative analysis decreases, the more importance is attached to formal properties of restrictive relatives and agreement patterns. However, the capacity of this analysis to account, more readily than others, for the relationship between agnate clefts and pseudo-clefts makes it more attractive for those whose standpoint is semantic, or functional in orientation.

4.2 HIGHLIGHTED ELEMENTS: FORM AND FUNCTION

As an aid towards a better understanding of cleft and pseudo-cleft constructions it is helpful to explore the range of syntactic classes and functions represented by their highlighted elements (that is, the complement of *be* in clefts and basic pseudo-clefts; the subject of *be* in reversed pseudo-clefts). While there is considerable overlap between clefts and pseudo-clefts in this respect, there are at the same time significant differences. Such differences need to be borne in mind in assessing analyses based on the claim (or commonly, the assumption) that clefts and pseudo-clefts are interchangeable. For instance Akmajian (1970), who posits a 'Cleft-Extraposition Rule' whereby clefts are derived transformationally from pseudo-clefts, claims (p.149):

There is a great deal of similarity between cleft and pseudo-cleft sentences. For example, the following two sentences:

- (1) a. The one who Nixon chose was Agnew.
b. It was Agnew who Nixon chose.

are synonymous, share the same presuppositions, answer the same questions, and in general they can be used interchangeably.

Furthermore, I shall argue, correlations occur between a number of these syntactic properties and informational/thematic properties of clefts and pseudo-clefts. Such correlations provide support for the principle, which is axiomatic to functional theories of language, that there is a systematic relationship between language structure and language function.

(i) Class

Seven classes of highlighted elements are represented in the corpus: noun phrase, prepositional phrase, finite clause, non-finite clause,

adverb phrase, adjective phrase, and zero. In addition to these, there are two further types of item that Delahunty (1982, 1984) claims to be possible in clefts, namely particles and quantifier phrases. He illustrates these using invented examples such as the following ((15a) = Delahunty 1.32.a; (15b), (15c) = Delahunty 1.43.c,d):

- (15) a. It wasn't on that he pulled his boots, it was off.
b. How long was it that we stayed under the water?
c. It's little we care about our mothers till they're gone.

Delahunty's class of quantifier phrases includes items from two of the classes used in the present classification: adverb phrases, as in (15b), and noun phrases, as in (15c). There is no example of the type represented by (15a) in the corpus. What this may suggest is that the particle type is of a highly specialized nature, restricted to contrastive contexts in which the speaker or writer is correcting or contradicting a previous assertion.

There are, furthermore, no Standard English examples of adjective phrases, as in (16) (= Delahunty 1.35a), in the corpus.

- (16) It was green that he painted the boat.

The only corpus example, (17g) below, is clearly dialectally marked. The adjective phrase type would seem to be, like the particle type, restricted to corrective or contradictory contexts.

Five of the seven classes are represented in significant numbers in clefts, and the remaining two by a handful of examples. The six examples of non-finite clauses (of which two are gerundial and four infinitival), could arguably be classified as noun phrases. A selection of examples is given below, with the order in which classes of highlighted elements are presented reflecting their frequency of occurrence in the corpus:

- (17) a. It was Mrs Kennedy who drew the crowds, said police.
(LOB A28.26)
b. # it was [ɔ:] {THROUGH} DÁVID that [ə:m] # -
INGRID met Don # (LL S.4.4,148-9)
c. If so, it must be that their God was more powerful than the Kikuyu's Ngai, (LOB K29,39-40)
d. It's when contractions are stronger, and more frequent, that she prefers to lie down. (LOB F32,52-3)
e. It is not so very long ago that Brahms met with bored incomprehension in Latin countries, that Bruckner and

Mahler were regarded as exclusively Teutonic, Fauré exclusively French, and Nielsen exclusively Scandinavian, while Anglo-Saxons and Scandinavians marvelled at their own particular appreciation of Sibelius, Delius or Vaughan Williams. (LOB A17,12-17)

f. It was indeed in order to obtain a localised wave function that the Wannier functions were first introduced. (LOB J05,180-2)

g. The girl beside her laughed again, her green eyes dancing in the firelight.

'Och, Mistress Paterson - you should be telling fortunes at a fair!' She teased. 'It's rich you'd soon be, with such fine lates for the asking!' (LOB N28,58)

Table 4.1 Clefts: class and function of highlighted elements in LL + LOB

Function Class	Subject	Direct object	Indirect object	Adjunct	Zero	Comp. prep	Total
NP	280	49		26		24	379 (50.4%)
PP			3	159		162	162 (21.5%)
ZERO					109		109 (14.5%)
FIN CL	6	2		43			51 (6.8%)
N-F CL	2			4			6 (0.8%)
ADV P				44			44 (5.9%)
ADJ P	1						1 (0.1%)
Total	289 (38.4%)	51 (6.8%)	3 (0.4%)	276 (36.7%)	109 (14.5%)	24 (3.2%)	752 (100%)

As indicated in Table 4.1 the commonest type of highlighted element in clefts, by a considerable margin, is the noun phrase, exemplified in (17a) (representing 50.4% of cases in the corpus). The second most frequently occurring highlighted element in clefts, representing 21.5% of cases in the corpus, is the prepositional phrase. It is Delahunty's (1984:75) view that cleft-highlighted prepositional phrases 'may range over the entire set of roles and functions assignable to that category'. With one exception (namely prepositional phrases as embedded qualifiers: compare 'It was the man with the moustache that was chosen' with 'It was with the moustache the man that was chosen') this claim is true. The two primary functions of

prepositional phrases are represented in the corpus: that of adjunct, associated with a wide array of semantic roles, and exemplified in (17b), and indirect object (as in (18) below):

(18) It is to these 'customers', the advisors of students, the creators of ambition, that we have to sell a new picture of the system, as it will be, a system in which Oxbridge will have a special but not predominant place. (LOB G61,102-5)

The third most common highlighted element in clefts, accounting for 14.5% of cases in the corpus (as exemplified in (17c)) is zero. With this category, discussed above in Section 3.2, the relative clause contains all items having an ideational function in the sentence, the cleaving serving to highlight non-ideational items relating to tense, modality, aspect and polarity.

Next in order of frequency are finite clauses, mostly adverbial clauses as in (17d), then adverb phrases as in (17c), non-finite clauses as in (17f), and the single adjective phrase in (17g).

Pseudo-clefts highlight a narrower range of syntactic classes than clefts in the corpus, with only the three classes of noun phrase, finite clause and non-finite clause being represented in significant numbers in the basic construction, and only the class of noun phrase in the reversed construction. The data for the present study bear out Prince's (1978:885) observation, based on a brief discussion of clefts and (basic) pseudo-clefts containing a range of highlighted constituents, that 'the only significant overlap concerns focused NPs'. However, as the discussion below indicates, further claims made by Prince in elaboration of her observation lack support from the present findings. A selection of examples is presented in (19), and frequencies for LL and LOB are given in Table 4.2.

- (19) a. # what he DOESN'T RÉALIZE # is that not EVERYBODY else # can work quite as hard as HE can # (LL S.1.5,207-9)
- b. Unfortunately, modern research tends to discredit the legend, claiming that *what really happened was a visit from a labour agent who attracted many loyal young men away to Bohemia, with the promise of good wages.* (LOB F35,57-9)
- c. # well what I was trying to DO # was to draw a line through THERE # (LL S.1.11,760-1)
- d. # I. I I think that - the WAY # that GÉOFFREY'S # been speaking about YÔU MARY # has been very MUCH # - of an IDOL # (LL S.6.5,962-6)

Table 4.2 Basic pseudo-clefts: class and function of highlighted elements in LL + LOB

Function Class	Subject	Direct object	Adjunct	Comp. prep.	Comp. subject	Comp. verb	Total
FIN CL	87	92	19	6	4		208 (44.8%)
NP	55	73	4	20	2		154 (33.2%)
N-F CL	10	8		1		78	97 (20.9%)
PP			4			1	5 (1.1%)
Total	152 (32.8%)	173 (37.3%)	27 (5.8%)	27 (5.8%)	6 (1.3%)	79 (17.0%)	464 (100%)

Most frequent as highlighted element in basic pseudo-clefts are finite clauses, mostly content clauses as in (19a). Basic pseudo-clefts do not display as wide a range of noun phrases (exemplified in (19b)) as clefts in the corpus. In her discussion of the common ground between clefts and pseudo-clefts involving highlighted noun phrases, Prince (1978:885) makes the assertion that:

Even within the set of NP foci, the two constructions differ in that an *it*-cleft may focus an animate or inanimate NP with equal facility, whereas (in Modern English) a *WH*-cleft may focus only an inanimate NP.

This claim is incorrect, unless we understand that by 'animate' and 'inanimate' Prince means 'human' and 'non-human' respectively. The following corpus example shows this to be so:

- (20) I sat down to figure out what they couldn't use for war, and what I came up with was birds. (LOB A26,82-3)

Next in order of frequency are non-finite clauses, mostly infinitive clauses serving as complements to *do*, as in (19c). There are a small number of highlighted prepositional phrases, as in (19d).

Reversed pseudo-clefts display an overwhelming preference for highlighted noun phrases, with other classes of highlighted elements accounting for only one per cent of cases in the corpus. Frequencies are presented in Table 4.3. The noun phrase class is dominated by demonstratives (see further Section 6.2.1 below), as in (21a). The other grammatical classes represented are non-finite clause (as in

(21b)), prepositional phrase (as in (21c)), adverb phrase (as in (21d)), and adjective phrase (as in (21e)).

- (21) a. # that's how it OPERATES # (LL S.2.2,749)
 b. Opening new branches is not however the only way in which we feel able to help with child care problems in the world at large. (LOB H25,67-8)
 c. # and up on the TOP of the [lae] on the left-hand SIDE # was where Ballylee CASTLE was # (LL S.1.14,165-6)
 d. 'Well, I never make it a practice to inquire into a man's personal or marital affairs, Constable, here is where I turn in.' (LOB N03,15)
 e. Exasperated is what Mr Roberts sounded. (LOB M04,57)

Table 4.3 Reversed pseudo-clefts: class and function of highlighted elements in LL + LOB

Function Class	Subject	Direct object	Adjunct	Comp. Prep.	Comp. Subject	Total
NP	79	218	193	57	16	563 (99.0%)
N-F CL	2					2 (0.4%)
PP			2			2 (0.4%)
ADJ P					1	1 (0.2%)
ADV P			1			1 (0.2%)
Total	81 (14.2%)	218 (38.3%)	196 (34.4%)	57 (10.0%)	17 (3.0%)	569 (100%)

What the discussion in this section indicates is that there are significant differences between cleft and pseudo-cleft constructions in the syntactic classes of highlighted elements chosen by each. Pseudo-clefts strongly favour nominal elements as a consequence of the means by which they highlight (namely, through direct equation, with the equated segments being reversible), giving rise to prominence in the theme of an ideational type (see further Section 5.2.3 below). The strength of the pressure to select a nominal element is evident with reversed pseudo-clefts, where the nominal demonstratives *this* and *that* are regularly selected as circumstantial identifiers at the expense of the demonstrative adjuncts *here*, *there* and *then* (for discussion of

the text-referential function of demonstrative *this* and *that* in reversed pseudo-clefts, see Section 6.2.1 below).

(By contrast with pseudo-clefts, clefts highlight virtually any item which is able to be thematized in the corresponding non-cleft. From a thematic perspective, the role of the cleft construction is to add to an already thematic item the further prominence deriving from predication in the context of the cleft construction (see further Section 5.2.3 below). Consider some typical examples:)

- (22) a. Despite chronological difficulties, it is they who have been suggested as the link between the arts of Central Asia and the Steppes. (LOB G45,178-9)
 b. It is with deep regret that we pay a last tribute to a great friend and colleague who has passed on. (LOB F16,203-4)
 c. # - and it's interesting to REMEMBER that # . it was AS a LIEUTENANT # that his FATHER {the Duke of EDINBURGH # } # entered the ABBEY # for HIS wedding # in nineteen forty-SEVEN # (LL S.10.6,450-4)

These are thematically agnate to 'They have been suggested . . .', 'With deep regret we pay . . .', and 'As a lieutenant his father . . .'

(The greater pressure that is exerted in pseudo-clefts than clefts for a nominal highlighted element to be chosen presumably derives in part from the fact that in pseudo-clefts the highlighted element is complementary to a nominal clause, and is therefore itself likely to be nominal. By contrast, in clefts, the highlighted element serves as complement to predicative *it*. The predicative use of non-referential *it* is one of a range of uses of the pronoun, giving it a syntactic flexibility that ties it less closely to the nominal category.)

For a large number of clefts with highlighted prepositional phrase there is no acceptable pseudo-cleft counterpart. The following pseudo-cleft, corresponding to (22b), is of questionable acceptability.

- (23) ? How we pay a last tribute to a great friend and colleague who has passed on is with deep regret.

The fact that clefts permit a wide range of prepositional phrases as highlighted element, but pseudo-clefts comparatively few, has been suggested by Halliday (unpub. MS, § 22.13) to follow from differences in the nature of the thematic highlighting across the two constructions ('textual' in the case of the cleft, and 'ideational' in the case of the pseudo-cleft). For many of the prepositional phrases

which are highlighted in clefts it is impossible to substitute a *wh*-item, suggesting that they cannot enter into an identifying relationship, as in a pseudo-cleft structure. They are, nevertheless, compatible with the textual form of highlighting of the cleft construction. Similarly compatible with the textual highlighting of the cleft construction, but incompatible with the ideational highlighting of pseudo-clefts, are the non-ideational highlighted items that I have designated as 'zero'.

One type of constituent that occurs in pseudo-clefts but not clefts is the non-finite clause that functions as complement to *do*.⁵ There is no cleft counterpart to the pseudo-cleft in (19c) above, for example (*'It was to draw a line through there that I was trying to do').

One category which, though possible as highlighted element in pseudo-clefts and clefts, displays markedly different frequencies of occurrence across the different constructions, is the finite clause. Finite clauses are the most common highlighted element in basic pseudo-clefts (see Table 4.2), but rank only fourth in clefts (see Table 4.1) and are unrepresented in reversed pseudo-clefts (see Table 4.3). It is not that reversed pseudo-clefts with a highlighted finite clause are ungrammatical. Rather, they contravene the principle of 'end-weight' in English (that is, the tendency for longer or more semantically weighted elements to be sentence-final: see for example Quirk *et al.* 1985:1361-2). Clefts with highlighted finite clause do slightly less violence to this principle since, even though the highlighted subordinate clause may outweigh the 'main clause' material which it precedes, it nevertheless occurs finally and with appropriate weight within the surface main clause in which it serves as complement to the subject *it*. Compare the relative acceptability of the cleft and reversed pseudo-cleft versions of the basic pseudo-cleft in (24b):

- (24) # the reason it was QUIET before # nineteen SIXTY-EIGHT # was because . you can ARGUE # is because . the British - didn't . didn't stir up the Northern [afor 5th am] the Ulster PROTESTANTS # (LL S.2.8,429-34)

('It was because the British didn't stir up the Ulster Protestants that it was quiet before.' 'Because the British didn't stir up the Ulster Protestants was the reason it was quiet before.')

(ii) *Function*

Seven syntactic functions are represented by the highlighted elements of clefts in the corpus. Of these, the subject and adjunct functions are dominant, together accounting for 75 per cent of instances.

Frequencies are given in Table 4.1 above. Examples of the seven functions are given below, with functions ordered according to their frequency of occurrence: subject (25a), adjunct (25b), zero (25c), direct object (25d), complement of preposition (25e), indirect object (25f), and subject complement (the only example is (17g) above).

- (25) a. It is apparently praise that is still so important to him that he lets his talent drown – (LOB C06,88–9)
 b. # I think it was through her INSPIRATION # that POSSIBLY # the Women's INSTITUTE {and things like THAT #} # really DEVELOPED # (LL S.12.6,995–9)
 c. # [ɔ:m] , it could well BE # that it's [ə: ʔ ə:] it's time to adjust it {UPWARDS} ANYWAY # (LL S.3.2,425–6)
 d. While this general picture of the way the farm was run will be of interest to the practical farmer, it is the economic aspect which the experiment was undertaken to test. (LOB E36,130–1)
 e. # – and it is the imagination and the MIND of MÂN . # that I'm INTERESTED in # (LL S.3.1,1123–4)
 f. Hattie glanced from one to the other in keen dismay. It was to Terence that she made her appeal. (LOB P13,171–2)

Five syntactic functions are represented by the highlighted elements of basic pseudo-clefts in the corpus. Of these the direct object and subject functions were dominant, together accounting for 70 per cent of instances. Frequencies for LL and LOB are given in Table 4.2 above. Examples of the five functions are given below, with functions ordered according to their frequency of occurrence: direct object (26a), subject (26b), complement of verb (26c), complement of preposition (26d) and adjunct (26e).

- (26) a. What the great masses of ordinary people in the world desire most of all is the certain prospect of peace for as long ahead as possible. (LOB B20,190–1)
 b. # but what is . IMPÖRTANT # and INTERESTING # is the political . MOVEMENT # . of our TIMES # (LL S.5.5,760–3)
 c. # what they did was to collect OPINIONS and VOICES # (LL S.5.3,243)
 d. What we can and do object to, however carefully 'landscaped' and however beautifully designed this power station may be, is the fact that we shall be able to see it from all parts of the Solent. (LOB E18,126–8)

- e. f. ÄND [ə] # . the reason it GÖT there # was because [ə:m]
 – my WIFE # [ə:m] .
 aud (laughter and applause)
 >f said to be [ð] I'd rather like you to get a Grecian ÜRN
 # (LL S.11.3,857–9)

Five syntactic functions are represented by the highlighted elements of reversed pseudo-clefts in the corpus. Again, two functions are dominant (direct object and adjunct), together accounting for 73% of instances. Frequencies are given in Table 4.3 above. Examples of the five functions are given below, with functions again ordered according to their frequency of occurrence: direct object (27a), adjunct (27b), subject (27c), complement of preposition (27d), and subject complement (27e).

- (27) a. Like the Grecian urn and beauty, that was all we knew or needed to know about Russia. (LOB G22,6–7)
 b. # and that's when it's going to BE # (LL S.2.10,203)
 c. # – that's what's FEMALE # (LL S.2.5,431)
 d. 'And that's all I'm interested in,' said Mr Harvey helplessly. (LOB A07,184)
 e. # which is what I regard my late holiday as having BEEN # (LL S.2.14,98)

As Tables 4.1, 4.2 and 4.3 indicate, the three main syntactic functions of highlighted elements in clefts and pseudo-clefts are subject, direct object and adjunct. (Although the 'zero' function associated with the non-identical highlighted elements in clefts ranks highly with them, it is not represented in pseudo-clefts, and is thus not central to the discussion of the 'main' functions below.)

Amongst the 'minor' functions, highlighted complements to a preposition are represented in all three constructions. The relative popularity of this function with reversed pseudo-clefts (10.0%: see Table 4.3) most likely has an informational explanation (notice that this function is considerably more popular in speech – 13.1% in LL – than it is in writing – 5.0% in LOB). The reversed pseudo-cleft construction is, of the three constructions, most consistently associated with a sentence-final preposition. (In clefts the preposition is most often positioned before the relative word. For further discussion see Section 3.3 above.) The reversed pseudo-cleft offers the speaker the opportunity of locating unmarked information focus on the verb, through the choice of a phrasal verb, and at the same time expressing

the further semantic features associated with the pseudo-cleft construction.

Highlighted complements to the subject are represented in pseudo-clefts (albeit in small numbers: see Tables 4.1, 4.2 and 4.3), but not (with the single exception of the dialectally marked example, (17g)) in clefts. Whatever explanation is offered for this restriction on Standard English will be parallel to that needed to account for the unacceptability of non-cleft sentences with a thematized predicative complement, as the following paired examples indicate:

- (28) a. ? The chairman he was.
 b. ? It was the chairman that he was.
 (29) a. ? Poor he was.
 b. ? It was poor that he was.

The function of verbal complement is associated, in pseudo-clefts, with highlighted non-finite clauses serving as complement to *do*. The fact that this function is not represented in reversed pseudo-clefts might be regarded as an accidental omission, but that it is not represented in clefts cannot be so explained. Compare the following examples:

- (30) a. What she was doing was drinking a sherry.
 b. Drinking a sherry was what she was doing.
 c. *It was drinking a sherry that she was doing.

There is no example of either a cleft or pseudo-cleft with a non-prepositional phrase indirect object in the corpus. As the variable acceptability of the following examples suggests, this is not surprising in the case of clefts and basic pseudo-clefts, but is perhaps an accidental omission in the case of reversed pseudo-clefts.

- (31) a. * It was her I sold the car.
 b. *? The one I sold the car was her.
 c. ? She was the one I sold the car.

Turning to the 'major' functions, one is struck by differences in their relative frequency across the three constructions. The relevant data, taken from Tables 4.1, 4.2 and 4.3, are presented in Table 4.4 below.

It was suggested above that the cleft construction functions, thematically, to imbue an already thematic item with 'textual' prominence. This being so we might expect the choice of highlighted functions in cleft constructions to reflect that of ordinary non-cleft declaratives.

Table 4.4 Relative frequencies of the major syntactic functions of highlighted elements in LL + LOB

Clefts		Basic pseudo-clefts		Reversed pseudo-clefts	
1	subject (38.3%)	object (37.3%)	object (38.3%)	object (38.3%)	
2	adjunct (36.7%)	subject (32.8%)	adjunct (34.4%)	adjunct (34.4%)	
3	object (6.8%)	adjunct (5.8%)	subject (14.2%)	subject (14.2%)	

And this is what we do find. In declarative clauses in general the subject is the unmarked theme, and is undoubtedly statistically most frequent. Of the marked themes, as Halliday (1985:45) observes, the adjunct is the most usual in ordinary declarative clauses. More strongly thematically marked, and correspondingly less frequent, are objects.

Brief mention might be made here of a further aspect of the textual prominence associated with the highlighted element in clefts. The notion of textual prominence has an intonational, as well as a thematic, basis. The cleft construction creates, through predication, a local clause-structure in which the intonation nucleus on the highlighted element is in its regular clause-final position. (The actual figures, for the 67 highlighted subjects in LL, are 48 (71.6%) with nuclear stress, and 19 (28.4%) without.) The cleft construction achieves maximum textual prominence, intonationally speaking, by giving intonational prominence most commonly to that syntactic function which is least likely to carry it in ordinary declaratives (the subject). Meanwhile, the function most likely to carry intonational prominence in ordinary declaratives (the object), carries it least commonly in clefts. Adjuncts fall between these two extremes. To illustrate the above points, consider the way in which the cleft in (32a), by contrast with various other constructions, ascribes maximum 'textual' prominence to the subject *John*.

- (32) a. It was JOHN who did it.
 b. The one who did it was JOHN.
 c. JOHN was the one who did it.
 d. JOHN did it.
 e. It was done by JOHN.

(32a) is an unmarked cleft in which the prominence associated with *John* derives from its intonational and informational salience, the

syntactic highlighting obtaining from predication, and its status as theme. (32b) is the corresponding basic pseudo-cleft, in which *John* is non-thematic. (32c) the reversed pseudo-cleft, is informationally and thematically close to (32a), but *John* here lacks the syntactic prominence that it enjoys in the cleft. There are at least two ways in which the non-cleft sentences in (32d) and (32e) differ from (32a), (32b) and (32c) collectively. Firstly the non-clefts lack the feature of exclusive identification. Secondly, in the non-clefts, information is not organized into two parts representing presupposed and non-presupposed material in such a way that *John* gains prominence by its exclusion from the presupposition. In both (32d) and (32e), *John* carries intonational and informational prominence, but only in (32d) does it have thematic status.

By contrast with clefts, as Table 4.4 shows, in basic pseudo-clefts the object function is strongly represented and the adjunct function weakly. Though ranked behind the object function, the subject is, as in clefts, frequently chosen. Together, the subject and object functions account for 70.1% of functions in basic pseudo-clefts, compared with only 45.1% of clefts. This preference for participant-related functions in basic pseudo-clefts is linked to the preference they display for nominal highlighted elements. These tendencies appear to override the pressure, evident in the case of clefts, to maintain the word order patterns of the corresponding non-cleft.

Reversed pseudo-clefts display a different order of preference of the three main functions from both basic pseudo-clefts and clefts. The primary form of prominence that the reversed pseudo-cleft construction gives to the highlighted element is thematic. Lacking the predicating structure of the cleft, we might expect the pseudo-cleft to select functions in an order which reflects an attempt to achieve maximum thematic prominence. In other words, we would expect the order to be the reverse of that with clefts: from that function which would be the most strongly marked theme in the corresponding non-cleft, to the least strongly marked. And this is in fact what we do find. Compare the following:

- (33) a. I need that.
 b. That I need.
 c. That's what I need.
 (34) a. I resigned (for) that (reason).
 b. (For) that (reason) I resigned.
 c. That's why I resigned.
 (35) a. That annoys me.
 b. That is what annoys me.

There would seem to be greater motivation for reverse-pseudo-clefting the (a) sentences in (33) and (34), than in (35). The former present as unmarked theme an item which, if thematic in the corresponding non-clefts ((33b) and (34b)), is marked. In the sentences in (35), by contrast, reverse-pseudo-clefting has no effect upon the thematic status of the subject *that*: it remains the unmarked theme. The ordering of reversed pseudo-clefts like (33c) before those like (34c) may be explained in terms of the greater degree of markedness of the theme in the corresponding thematically fronted non-cleft, (33b).

4.3 SEMANTIC PROPERTIES

Attention has already been paid (in Chapter 1 above) to some of the semantic aspects of cleft and pseudo-cleft constructions, in the course of defining the class. In this section I shall examine the identifying structure of clefts and pseudo-clefts, the exclusiveness implicature associated with this structure, and the existential presupposition conveyed by the relative clause of the constructions.

(i) Identification

In Section 1.2 above it was argued that clefts and pseudo-clefts are a form of identifying construction, expressing a relationship between an element that is to be identified (the 'identified') and an element that identifies it (the 'identifier'). The identified corresponds to what we have called the relative clause, the identifier to the highlighted element. In pseudo-clefts, as in non-cleft identifying sentences, the identified-identifier sequence is reversible, giving rise to the two types of pseudo-cleft that I have labelled 'basic' and 'reversed'. As noted in Section 1.2 above, it is the property of reversibility that represents the fundamental grammatical difference between the modes of identification (in which there are two 'participants') and attribution (in which there is only one 'participant'). By contrast with pseudo-clefts, in clefts the identifier-identified sequence is not reversible, as a result of the structural device of predication, which (as discussed in more detail in Section 5.2.3 below) confers upon the identifier/highlighted element a special form of prominence.

There is a further dimension to identifying clauses which must be mentioned here, that characterized by Halliday (1967a) using the terms 'value' and 'variable' (corresponding to 'value' and 'token' respectively in Halliday (1985)). Whereas the token specifies the form of the target element, indicating how it is recognized, the value

specifies its function, indicating how it is valued. Halliday (1985:115-16) defines the relationship between the functions of token-value and identifier-identified as follows:

In any identifying clause, one element will be the Value (meaning, referent, function, status, role) and the other will be the Token (sign, name, form, holder, occupant). These functions are then conflated with those of Identified and Identifier; and the conflation can go either way. Either the Token or the Value can serve as the identifying element (the Identifier).

Halliday uses the term 'decoding equatives' for constructions in which the identifier is mapped on to the value (that is, in which the identification proceeds by decoding). 'Encoding equatives' are those in which the identifier is mapped on to the token (that is, in which the identification proceeds by encoding).

Halliday (1967a) has argued convincingly that both pseudo-clefts and clefts are encoding equative constructions. He rejects decoding constructions from the class of pseudo-clefts on the grounds that they are not agnate to a (thematically congruous) non-equative clause. For example the presupposed question for a 'decoding equative' such as *What they sell are bargains* is also an equative, namely *Are what they sell bargains?*, rather than the non-equative *What do they sell?* Halliday concludes that (p.231):

The feature 'identifying' ['pseudo-cleft', in our terminology] may then be said to assign a structure merely in terms of identified and identifier, since in any clause whose equative structure was derived from this option the assignment of the functions value-variable would be predictable therefrom: the identified is always the value and the identifier the variable.

In clefts too, argues Halliday (1967a:238-9), the identifier always functions as the token:

It seems to be impossible to predicate the value in an equative relation, so that it's *the leader that's John* can only be interpreted as 'this is how John can be recognized'. This is related to the alignment of functions in identifying clauses, and means that the identified element can never be predicated; but this is predictable from the fact that what is predicated has the role of an identifier within the predication.

The corpus data used in the present study do not yield any counter-examples to Halliday's claims regarding the token-value structure of

pseudo-clefts and clefts. Nevertheless the question is one that certainly warrants further study, and in a degree of detail that is beyond the scope of this book.⁶

(ii) *Exclusiveness implicature*

A semantic feature of clefts and pseudo-clefts, which derives from their equative form (and which was introduced in Section 3.1.2 above), is that of exclusiveness. Halliday (1985:43) explains the meaning of this feature as follows:

the meaning of *what the duke gave my aunt was that teapot* is something like 'I am going to tell you about the duke's gift to my aunt: it was that teapot - and nothing else'. Contrast this with *the duke gave my aunt that teapot*, where the meaning is 'I am going to tell you something about the duke: he gave my aunt that teapot' (with no implication that he did not do other things as well).

Following Halvorsen (1978), I shall regard the exclusiveness feature as a 'conventional implicature'. The term 'implicature' refers to what speakers can imply, as opposed to what they literally say.⁷ Conventional implicatures may be distinguished from 'conversational implicatures'. Conventional implicatures are determined by the conventional meanings of linguistic expressions. In the following examples borrowed from Halvorsen (1978) (who in turn borrowed them from Karttunen and Peters (1975)), (36b) is an assertion (or, entailment) of (36a), whereas (36c) is not directly asserted (or, entailed) but rather conventionally implicated by (36a).

- (36) a. John managed to write a paper to present at the conference.
b. John wrote a paper to present at the conference.
c. It is difficult to write a paper to present at the conference.

Whereas conventional implicatures are part of the linguistic system, arising whenever the linguistic expression with which they are associated is used (*manage to*, in the case of (36a)), conversational implicatures fall within the realm of pragmatics. Conversational implicatures are determined by the linguistic and non-linguistic context in which an expression is used. They arise from the interplay of what is said with one or more principles of conversation (Grice's 'conversational maxims'). Grice (1975) discusses the discourse fragment in (37).

- (37) A Smith doesn't seem to have a girl-friend these days.
B He has been paying a lot of visits to New York lately.

claiming that unless B's contribution in (37) is assumed to carry the conversational implicature that 'Smith has, or may have, a girl-friend

in New York', there will be a violation of the maxim of relation. According to this maxim, the contribution of each conversational participant 'must be appropriate to the immediate needs at each stage of the transaction'.

A number of writers use the term 'implicature' in a way that is little different from the notion of presupposition. For example in Halvorsen's, and Karttunen and Peters' conception, conventional implicature appears to differ from presupposition only in that it covers a slightly wider range of phenomena, referring to almost any inference that is conventionally licensed by a linguistic expression. Apart from this, conventional implicatures, like old-fashioned presuppositions, but unlike entailments, are understood to be preserved under negation and questioning. Indeed, in their conception of conventional implicature as little different from presupposition, Halvorsen, and Karttunen and Peters, depart from Grice's original notion of conventional implicature, which was intended to be distinct from presupposition (in that it was not carried by what was actually 'said'). There are other writers who have extended the notion of presupposition in order to incorporate, amongst other things, the Gricean concept of implicature. For example Gilhotra (1984) distinguishes three types of presupposition, 'logical', 'interpersonal' and 'textual', and discusses implicature as a subcategory of the textual category.

Halvorsen (1978:15-16) notes some of the difficulties involved in formulating the exclusiveness implicature of clefts and pseudo-clefts. Consider the following sentences (= Halvorsen (59a), (66), and (68):

- (38) a. It was John that Mary kissed.
b. John was the only person that Mary kissed.
c. Mary kissed only one person.

(38b) is not an adequate rendering of the exclusiveness implicature of (38a), because it is not an implicature of the *yes/no* question corresponding to (38a), namely *Was it John that Mary kissed?* (given that questions do not conventionally implicate their answers). (38c) avoids this problem, but raises others. Because the meaning of the exclusiveness implicature is exhaustive, rather than unique, (38c) would not be suitable as the exclusiveness implicature for *h* was *John and Bill that Mary kissed*. Higgins (1979:154) in fact compares pseudo-clefts to lists, noting the similarity between (39a) and (39b) (= Higgins (49a) and (49b) respectively):

- (39) a. What I bought was a punnet of strawberries and a pint of clotted cream.
b. I bought the following things: a punnet of strawberries and a pint of clotted cream.

This formulation could equally be applied to clefts, such that in both types of construction the exclusiveness implicature is taken to be an exhaustive listing of the entities which satisfy the identified clause. Halvorsen argues, however, that Higgins's characterization faces problems when, once again, questions are considered. The question *Was it John that Mary kissed?*, argues Halvorsen (1978:16) 'does not conventionally implicate that the list containing the single name *John* is an exhaustive listing of the persons Mary kissed. If it had been so, the question would have been pointless.' Halvorsen himself attempts to formalize the exclusiveness implicature using a 'model theoretically interpreted fragment' of a Montague grammar. This fragment uses explicit translation rules in an attempt to account for the implicational differences between pseudo-clefts and clefts on the one hand, and their non-cleft counterparts on the other, along with their truth conditional similarities. Halvorsen characterizes his treatment of the exclusiveness implicature of clefts and pseudo-clefts, which is the most comprehensive that has so far been presented, as representing but 'a first stab' (p.16) at the question, and admits that it has many problems (p.95). Clearly, further study of this implicature is needed.

(iii) *Existential presupposition*

Most writers who have commented upon the logical aspects of cleft and pseudo-cleft constructions have noted that they are truth functionally equivalent to their non-cleft counterparts, but differ from them in terms of the presence of an existential presupposition associated with their relative clauses.⁸ Consider again the sentences in (1) in Chapter 1 (repeated here for convenience as (40)). (40b) and (40c), but not (40a), would be said to presuppose that 'Tom offered Sue something'.

- (40) a. Tom offered Sue a sherry.
b. What Tom offered Sue was a sherry.
c. It was a sherry that Tom offered Sue.

The criterion for presuppositionhood used is that typically invoked in transformational grammar: 'a sentence *S* presupposes a sentence *S'*' just in case *S* logically implies *S'*' and the negation of *S*, $\sim S$, also

logically implies S' (Keenan 1971:45). So, it would be argued, in order to prove that (40b) and (40c) presuppose that 'Tom offered Sue something', whereas (40a) does not, we have only to consider the negated versions of these sentences: (41b) and (41c) below, but not (41a), presuppose that 'Tom offered Sue something'.

- (41) a. Tom didn't offer Sue a sherry.
b. What Tom offered Sue wasn't a sherry.
c. It wasn't a sherry that Tom offered Sue.

Halvorsen (1978) uses the notion of conventional implicature, which as we have seen he uses in a way that is little different from the traditional notion of presupposition, to express the existential presuppositional differences (as well as the exclusiveness difference noted above) between sentences like (40a) and (40b-c), and (41a) and (41b-c). Halvorsen claims that 'the existential implicature of a cleft sentence is determined solely by the form of the clefted clause' (p.18). He compares the following sentences:

	ASSERTION	IMPLICATION
(42) a. It was John that Mary kissed.	+	+
b. It wasn't John that Mary kissed.	-	+
c. It was John that Mary didn't kiss.	-	-
d. It wasn't John that Mary didn't kiss.	+	-

The sentences in (42) have four different 'meanings', because those that make identical assertions differ in their implicatures, and those that share implicatures differ in their assertions (as indicated in the feature matrix presented above, where '+' means 'positive' and '-' means 'negative'). Sentence (42a), claims Halvorsen, asserts that 'Mary kissed John' and implicates that 'There is somebody that Mary kissed'; (42b) asserts that 'Mary did not kiss John' and implicates that 'There is somebody that Mary kissed'; (42c) asserts that 'Mary did not kiss John' and implicates that 'There is somebody that Mary didn't kiss'; and (42d) asserts (though Halvorsen acknowledges that this may be controversial) that 'Mary kissed John' and implicates that 'There is somebody that Mary didn't kiss'.

Halvorsen rejects the example in (43) (= Halvorsen (64a)), on the grounds that the parenthetical clause transgresses the existential implicature associated with the cleft.

- (43) It wasn't John that Mary kissed, (she didn't kiss anybody).

Atlas and Levinson (1981), however, claim that the sequence in (43) is quite acceptable. They propose that, while the implicatures which Halvorsen identifies for sentences (42a) and (42c) are appropriately named, those which Halvorsen associates with (42b) and (42d) are in fact assertions (or, in their terminology, 'entailments'). The validity of Atlas and Levinson's proposal is, however, undermined by their faulty interpretation of the sequence in (43). The final clause is not a contradiction of an assertion made in the cleft: it is a contradiction of a presupposition that the speaker is attributing to his/her interlocutor.

Delahunty (1982) is, like Atlas and Levinson (1981), motivated by a desire to reinterpret the phenomenon of presupposition in terms of entailment along with some general principles of conversational inference. Delahunty says (p.189):

It was widely held until recently that clefts were a paradigm case of presupposition. But the considerable evidence amassed over the last several years argues that the relation glossed a logical or semantic presupposition can be accounted for by the independently necessary relationship of *entailment* and some non-truth-conditional component of the grammar.

Delahunty compares sentences (44a), (44b) and (44c) (= Delahunty (1), (2), (3) respectively), arguing that (44b) is not presupposed by (44a) or (44c).

- (44) a. It was an apple that John ate.
b. John ate something.
c. It was not an apple that John ate.

Delahunty's case hinges crucially, however, on an example of dubious relevance, namely:

- (45) If it wasn't an apple that John ate then John ate nothing.

(45) could only occur coherently in discourse as a challenge to a previous speaker's presupposition, and thus does not constitute counter-evidence to the claim of presuppositionhood.

I shall maintain that the existential presupposition that has traditionally been held to differentiate clefts and pseudo-clefts from their non-cleft counterparts is in fact appropriately named. Those who wish to dispense with the notion of presupposition (Atlas and Levinson, and Delahunty) have misinterpreted the data which they

adduce as evidence. Furthermore it is questionable whether the notion of presupposition can be subsumed under that of conventional implicature, as in Halvorsen's treatment. The exhaustiveness implicature and the existential presupposition are different in kind. Presuppositions are expressible as propositions whose formulation is generally retrievable from the form of the sentence. This is clearly so with clefts and pseudo-clefts, where the presupposition is an aspect of the core structure of the sentence. However it is also true of a sentence such as *The King of France is bald*, where the presupposition, 'there is a king of France', is retrievable from the noun phrase, *the King of France*. It is their status as propositions that gives presuppositions the potential to be questioned or contradicted in discourse, as in (44) and (45). By contrast with presuppositions, conventional implicatures are more 'subliminal'. Although they are expressible as propositions, the form of such a proposition is not determined in the same way by the structure of the sentence. As such they are less readily cancellable in discourse. Compare the following:

- (46) A It was Tom you saw. B No, I didn't see anyone.
 (47) ?A It was Tom you saw. B No, I saw other people as well.

In (46) Speaker B is able to challenge, quite felicitously, the existential presupposition conveyed by A's cleft ('you saw someone'). By contrast in (47) the proposition expressing the exclusiveness implicature conveyed by Speaker A's cleft ('Tom was the only person you saw') is less directly associated with the form of the sentence, and is thus not able to be directly challenged with equal acceptability.

The existential presupposition of cleft and pseudo-cleft constructions is mapped onto their identified element. As such, it directs the addressee's attention towards the element that 'satisfies the definition' expressed in the presupposition, namely the highlighted element. Also typically mapped onto this identified element in clefts and pseudo-clefts, is the function of givenness. The concepts of givenness and presupposition are closely related, both involving information which the speaker presents as known, or accessible, to the addressee, but they are nevertheless distinguishable. The differences are discussed in Section 5.3.3 below.

5 Communicative meanings

In this chapter I shall explore the notion of 'communicative meanings' and discuss their encoding in cleft and pseudo-cleft constructions. Because I shall be drawing extensively on the work of Halliday, the chapter will begin with an outline of his work on the textual component of language and its place within his functional model of grammar. This will be followed by a detailed examination (with particular attention to the work of Halliday) of the various conceptions that are to be found in the literature of the linguistic notions of information and theme.

5.1 HALLIDAY'S 'TEXTUAL' COMPONENT

Halliday's work on textual meanings cannot be considered in isolation from his functional approach to language, his view (1970c:323) that:

the structure of language as a whole has been built up in such a way that it reflects the demands that are made on language and the functions it is required to serve.

According to Halliday there is a basic functional organization inherent in the grammatical system, and determining the structure of the clause. Three 'macrofunctions', or generalized uses, of language may be distinguished, each corresponding to a major semantic system. These are:

- (i) the *ideational* function, which encodes our experience of the world; it is concerned with the expression of content. In the clause this function specifies the available options in experiential meaning and determines the nature of their structural realizations within the system of transitivity.

- (ii) the *interpersonal* function, which is concerned with the expression of social and personal relations, including all forms of the speaker's intrusion into the speech situation. In the clause the interpersonal function is represented by mood and modality.
- (iii) the *textual* function, which 'fills the requirement that language should be operationally relevant - that it should have a texture, in real contexts of situation, that distinguishes a living message from a mere entry in a grammar or dictionary' (1973:42). In the clause this function is represented by the system of theme and, in the tone unit, by that of information.

It is the textual function with which we shall be concerned in this chapter. Insofar as it is concerned with the creation of text this function is not only internal to language, but also instrumental to the other two. As Halliday says (1973:107):

It is through this function that language makes links with itself and with the situation; and discourse becomes possible, because the speaker or writer can produce a text and the listener or reader can recognize one.

The textual function specifies both inter-sentence relations (that is, the resources of cohesion) and intra-sentence relations (that is, the structural resources which Halliday (1967a:200) has referred to collectively as the 'theme system complex'). Here 'theme' is used generically and embraces the text-forming principles referred to in Prague School theory as 'functional sentence perspective', or 'FSP'. The systems encompassed by this complex yield an extensive paradigm of thematic options for the grammar of the English clause, whereby the speaker 'organizes his act of communication as a component of a discourse' (Halliday 1967a:241).

Two systems are said to operate in the case of every utterance: 'information' (the organization of discourse into a linear succession of intonationally realized units, each internally structured in terms of 'new' and 'given' information) and 'theme' (or 'thematization': the organization of the clause into a 'theme' - realized in English by the initial section of the clause - and 'rheme'). In addition there are four systems, each of which may combine with information and thematization to produce a subtle variety of communicative effects. The first two, 'identification' and 'predication', involve the organization of the clause into the equative structures that are referred to respectively in this book as pseudo-clefts and clefts. The remaining two, about which I shall subsequently have no more to say, are 'substitution' and

'reference'. Substitution (referred to as 'right dislocation' in formal grammars) reverses the normal sequence of theme-rheme by first introducing the theme as a cataphoric pronoun and then displacing it to the end of the clause (for example *They played some fine matches, the Wallabies; It's capable of 200km/h, that car*). Reference (known as 'left dislocation' in formal grammars) gives the theme prominence by picking it up later in the clause by an anaphoric pronoun (for example *Mr Hawke, he won with a reduced majority*). The anaphoric pronoun both isolates the theme and gives a separate specification of its transitivity role in the clause.

The sections which follow contain a more detailed account of Halliday's systems of information and thematization, with particular attention paid to those aspects of his account that are relevant to the present study of clefts and pseudo-clefts. Various criticisms that have been made of Halliday's model are discussed, and some alternative accounts of information and theme are considered.

5.2 THEME

The notion of theme, which has been used in Prague School thinking about language since at least the 1930s, has increasingly attracted the attention of British and American linguists over the past few decades as part of the growing interest in discourse as an object of study. In the sections which follow I shall describe two major approaches to the definition of theme, and provide an outline of Halliday's work on theme (since his name is probably most closely associated with the concept in linguistics today). This will be followed by a brief discussion of Fries's view that the choice of themes in a text constitutes its 'method of development'.

5.2.1 'Combiners' and 'separators'

Of the various approaches that have been taken to the definition of theme, most can be traced back to the definition proposed by Mathesius in 1939: 'that which is known or at least obvious in the given situation, and from which the speaker proceeds' (1939; translated by Firbas 1964:268). The first part of this formulation refers to givenness, while the second leads to the idea of theme in Trávníček's (1962) and Halliday's sense. Mathesius and the linguists who, like him, use two concepts to define theme, are labelled 'combiners' by Fries (1981). Fries says of the combiners (pp.1-2):

While they have all made various important refinements on the original Mathesian definition, the basic point remains that in their view the thematic content or topic of a sentence is determined to a great degree by how the information expressed in that sentence relates to the information already available in the linguistic and non-verbal contexts. That is, the theme of a sentence must be known information and one can only tell what is or is not known information in a particular sentence by looking at how the information in that sentence relates to the information available in the context.

The definitions of 'theme' and 'rheme' offered by Firbas (1957) are representative of the 'combining' approach. Firbas defines theme, or thematic elements, as (p.72):

elements which convey something already known or something that may be taken for granted, in other words, those elements that may be inferred from the verbal or from the situational context.

The other major approach to the definition of theme is characterized by Fries as the 'separating' approach. Advocates of this approach ('separators'), notably Halliday (1967a, 1970b, 1985), Trávníček (1962), and linguists of the Systemic school, distinguish between the concepts referred to in the two components of the Mathesian definition. Separators regard given information and theme as functionally distinct concepts which enter into different structural configurations. They claim, furthermore, that in English theme is realized by the element or elements occupying initial position in the clause. This claim gives rise to a major difference of opinion with the combiners. Since in the combining approach theme is defined in terms of known, or given, information, and since given information may occur anywhere within the clause or be absent from it, it follows that theme in this approach will not necessarily be clause initial in English and that some clauses will be regarded as themeless. Firbas (1966), for instance, argues that although there is a strong tendency in Modern English to express the theme of the sentence by means of the grammatical subject, non-thematic subjects do occur. In a sentence such as,

- (1) A girl came into the room.

argues Firbas, the subject noun phrase *a girl*, being a 'first-mention form' (that is, contextually independent, accompanied by an indefinite determiner) and being linked up with a verb conveying the

notion of existence or emergence, will express the rheme. In the separating approach the sequence rheme-theme is not regarded as possible in English because word order makes an independent contribution to the meaning of the clause, signalling the 'point of departure' independently of whatever given information there may be in the clause. In the combining approach the contribution of word order in the realization of theme in English is either ignored altogether (by van Dijk (1977) for instance) or treated as contributing to the same concept as the given-new distinction. The latter position is reflected in Firbas's concept of 'communicative dynamism' or 'CD', the contribution made by different sentence elements towards the development of the communication. As Firbas puts it (1964:270):

By the degree of CD carried by a sentence element we understand the extent to which the sentence element contributes to the development of the communication, to which it 'pushes the communication forward', as it were. It is obvious that elements conveying new, unknown information show higher degrees of CD than elements conveying known information.

5.2.2 Halliday on theme

Theme is defined informally by Halliday as the point of departure of the clause, 'the peg on which the sentence is hung' (1970b:161). The rheme is the remainder of the message, and develops the theme.) Halliday warns that while theme may be analysed as the element(s) assigned initial position in the clause, it is not in fact defined in this way. The concept of theme is a functional one in the Halliday model, and the definition of it as the point of departure for the clause relates to this function. First position is, argues Halliday, simply the means by which this function is realized in English. There are, he observes, other languages which have a category of theme where the means of realization is not position in the clause. (For instance in Japanese the means of realization is association with a particle.)

In cases of unmarked information structure (see below) the theme will be associated with given information via the 'good reason' principle. That is to say that a speaker's theme will, if there is no good reason for doing otherwise, be chosen from within the given (from what is, in other words, already 'in the air'). (However, because there is nothing preventing the speaker from choosing anything s/he likes as a point of departure, be it or be it not associated with given information, the concepts of theme and given must be separated as

independent options. Here is Halliday's (1967a:212) oft-quoted summary of the difference:

while 'given' means 'what you were talking about' (or 'what I was talking about before'), 'theme' means 'what I am talking about' (or 'what I am talking about now'), and, as any student of rhetoric knows, the two do not necessarily coincide.)

Thus, whereas the system of information structures communication in a way that relates it to material which is recoverable by the addressee, the theme system structures it independently of such material.

It is in the intersection of the systems of theme and mood that Halliday finds the most effective means of explaining the meaning of the choice of theme. The unmarked theme of any clause proclaims the status of the clause as a proposition. In independent clauses this principle is strongest with the non-declarative (or 'marked') moods. In interrogative clauses, for example, the typical (unmarked) theme is the element bearing the meaning 'I want to know something': the *wh*-element in a *wh*-interrogative, and the finite verb in a *yes/no* interrogative. The selection of theme is a meaningful choice within the clause, however, and the speaker may therefore opt to override the enunciation of mood by selecting a marked theme. Marked theme is a foregrounding of the speaker's point of departure and, as such, tends to appear as a separate information unit. Halliday explains this in terms of a sort of 'quantum of thematic potential' associated with the structure of the message in ongoing spontaneous discourse. Marked themes use up much of this potential (that is, leave little room for further thematic choice) and the result is a clear division into two parts within the clause. Unmarked themes, on the other hand, may not exhaust the thematic potential of a clause. This is particularly the case with coordinating and subordinating conjunctions, which occur obligatorily in initial position in the clause. In cases where thematic prominence is not sharply discontinuous with rheme, but diminishes gradually, it may be difficult to determine just how much material is thematic. For this reason Halliday (unpub. MS: § 22.17) adopts the following general principle:

if the first element has an experiential function, that element alone is thematic; if not, then the Theme extends up to (including) the first experiential element.

It follows that a sentence such as the following (from Halliday 1985:55),

- (2) On the other hand maybe on a weekday it would be less crowded.

has a multiple theme. The first two elements, *on the other hand* and *maybe* function respectively as textual and interpersonal adjuncts. These types of adjunct are particularly prone to being selected as theme: they have no role in the transitivity structure and thus represent an attenuated foregrounding. Accordingly, there is still some thematic potential left over, allowing for the further choice of the marked experiential (or 'topical') theme, *on a weekday*. The validity of Halliday's definition of theme as the 'point of departure of the sentence as message' is supported by research conducted by Fries (1981). Fries demonstrates convincingly that the choice of theme, clause by clause, is by no means a haphazard matter, but plays a fundamental role in the way discourse is organized. The choice of successive themes in a text constitutes what is often referred to as the 'method of development' of that text. Fries shows that narrative texts, which typically relate sequences of events involving a common character or set of characters, tend to exhibit sequences of clauses in which each clause has a similar theme to the previous one (reference to a character or characters, or to sequence in time, and so forth). By contrast in scholarly articles the method of development tends to be a 'step by step' one. Complex arguments are presented in which each successive idea is an expansion of, or dependent upon, an idea in a previous sentence. Typically in this genre the theme in each successive sentence picks up an idea presented in the rheme of the previous sentence. Further work on thematic (and topical) progression in texts has been conducted by, amongst others, Givón (1983) and Martin (forthcoming).

5.2.3 Theme in pseudo-clefts and clefts

Halliday (1985:43) has suggested that pseudo-cleft constructions, which he calls 'thematic equatives' in this publication, have evolved in the language as thematic bracketing devices which enable a speaker to select almost any element or group of elements as his theme:

A thematic equative (which is sometimes called a 'pseudo-cleft sentence' in formal grammar) is simply an identifying clause with a thematic nominalization in it. Its function is to express the Theme-Rheme structure in such a way as to allow for the Theme to consist

of any subset of the elements of the clause. This is the explanation for the existence of clauses of this type: they have evolved, in English, as a thematic resource, enabling the message to be structured in whatever way the speaker or writer wants.

The pseudo-cleft construction enables subsets of elements to be grouped into two parts in an almost unlimited number of ways. Thus sentence (1a) in Chapter 1 (reproduced below as (3a)) might be grouped not simply as (1b) (= 3b), but also as in (3c-g), to name but a few possibilities.

- (3) a. Tom offered Sue a sherry.
 b. What Tom offered Sue was a sherry.
 c. The one who offered Sue a sherry was Tom.
 d. What Tom did was offer Sue a sherry.
 e. What Tom did with the sherry was offer it to Sue.
 f. What happened was that Tom offered Sue a sherry.
 g. What happened to Sue was that Tom offered her a sherry.

Typically, in basic pseudo-clefts, only one element is rhematic. In cases where more than one element occurs in the rheme, there is a general rule that one of these elements must be the verb (as in (3d-g)).

As is the case with equative constructions in general, the sequence of elements functioning as identified and identifier is free; that is, either may occur in first position as the theme. Thus in reversed pseudo-clefts such as *Tom was the one who offered Sue a sherry* and *Offer Sue a sherry was what Tom did*, the themes are respectively *Tom* and *offer Sue a sherry*. The pseudo-cleft, then, provides a discontinuous theme-rheme structure (strictly speaking the copula is outside the structure) with a clearly defined constituent structure. This discontinuity is reflected in the information structure where, in the typical instance (see Chapter 7 for corpus statistics), a basic pseudo-cleft is realized as two information units, with the boundary falling at the end of the theme. If the pseudo-cleft is realized as a single information unit the focus typically falls, as in equative clauses generally, on the identifier/highlighted element:

- (4) # the one who offered Sue a sherry was TOM #

However, as Halliday (1967a:226) has observed, it is not possible simply to interpret the identified-identifier structure of (basic) pseudo-clefts as corresponding to given-new information. The two

are independently variable; the association of new information with the highlighted element may be overridden by marked focus, as in:

- (5) a. Tom offered Sue a sherry.
 # NO # the one who POURED Sue a sherry was TOM # ('instead')
 b. Tom offered Sue a sherry.
 # and he was the one who POURED it for her # ('in addition')

One of the most common types of pseudo-cleft, the reversed type with demonstrative as highlighted element, normally has information focus on the relative clause (see Chapter 7 for corpus statistics):

- (6) # that's what he WANTS #

Like pseudo-clefts, clefts are equative structures, but in this case the identifier/highlighted element always precedes the identified/relative clause, and is thus thematic. (In clefts any element having a representational (or 'deep structure') function may appear as the theme.) In Halliday's earlier writings (for example Halliday 1967a) clefts were analysed as exhibiting a theme-rheme structure which was closely associated with their discontinuous syntactic structure. Thus in sentence (1c) in Chapter 1, reproduced as (7) below,

- (7) It was a sherry that Tom offered Sue.

a sherry would have been the theme, and *it . . . that Tom offered Sue* the rheme. In Halliday's later work (1985:60) a dual thematic analysis is given for clefts. The first level is said to represent the 'local, congruent thematic structure' of the two clauses in the construction. The second level reflects Halliday's treatment of cleft constructions as grammatical metaphors (the congruent version being the corresponding non-cleft version), where the choice of metaphorical form is motivated by thematic considerations. According to this version (7) would be analysed as follows:

- (8)
- | | | | | |
|-----|-------|--------------|----------|-------------|
| | It | was a sherry | that Tom | offered Sue |
| i. | Theme | Rheme | Theme | Rheme |
| ii. | Theme | | Rheme | |

In the typical instance (see Chapter 7 for corpus statistics) information locus falls on the theme, yielding a combination which, in non-cleft sentences, is marked. However, in clefts the theme/new combination is unmarked: the construction creates, through predication, a local structure – the superordinate clause – in which information focus is in its unmarked place, at the end. The interaction of theme and information in clefts provides one clue to their popularity in writing (see further Chapter 7): since stress is not marked in writing the construction serves to direct the reader into the intended interpretation of the information structure.

There is a broad thematic similarity between clefts and reversed pseudo-clefts, in their structural mapping of theme on to identifier, and unmarked mapping of these functions on to new information. Nevertheless, it has been argued by Halliday (1967a; unpub. MS) that there is a difference in the kind of theme highlighting involved in the two cases. The pseudo-cleft isolates, and thereby gives prominence to, its theme by representing it as one of the two members participating in an equative relationship. The type of prominence carried by the theme is essentially cognitive, or ideational, in flavour; that is, the function of the theme is to select a particular participant in the relational process. In clefts, on the other hand, the theme is predicated; in other words, it is presented structurally as a predicative complement. The elements realizing the theme and rheme are not reversible: the emphasis is on 'picking out' (or, predicating) one part of the structure. Thus the theme in clefts embodies a prominence that is essentially textual in flavour.

What evidence is there to support this contention that there is a semantic difference in the types of thematic prominence carried by the highlighted element in reversed pseudo-clefts and clefts? One of the syntactic differences between pseudo-clefts and clefts that was observed in Section 4.2 above was that the latter, but not the former, allow a wide range of prepositional phrases to occur as highlighted element, and that many of these are of a type that cannot enter an identifying relationship of the type in pseudo-clefts. On the other hand there is nothing preventing their being highlighted as themes in a structure where the theme is predicated (that is, in clefts).

Halliday notes that the difference between ideational and textual prominence stands out more clearly if the superordinate clause of the cleft and corresponding reversed pseudo-cleft is either negated or converted into the interrogative, or both. In the a. sentences of (9), (10) and (11) below the theme *Tom* carries ideational prominence as a participant in the process of offering. In the b. sentences the theme

Tom carries textual prominence as predicated topical theme: it is *Tom's* identity that is highlighted, rather than his participation in the process. The interrogative in each case questions the identity of the theme and not the ideational content of the message.)

- (9) a. Tom wasn't the one who offered Sue a sherry.
b. It wasn't Tom who offered Sue a sherry.
- (10) a. Was Tom the one who offered Sue a sherry?
b. Was it Tom who offered Sue a sherry?
- (11) a. Wasn't Tom the one who offered Sue a sherry?
b. Wasn't it Tom who offered Sue a sherry?

Finally, Halliday notes the occurrence of marked thematic clefts of the type,

- (12) Then it was I began to repent I had not given ear to the wholesome advice Cecil would have given me concerning the secret conduct of the Earl of Essex. (LOB K20,33-5)

whose function is probably to assert explicitly the thematic status of the highlighted element. The marked thematic construction avoids what might be a temporary ambiguity caused by the initial *it*. Whereas the *it* of the cleft construction is only weakly thematic, possessing thematic status only at a local level, anaphoric *it*, when sentence-initial, will always be thematic.

5.3 INFORMATION

The notion of 'given' and 'new' information is used in the linguistic study of texts by such writers as Halliday, Chafe, and members of the Prague School including Daneš and Firbas. Generally speaking, the terms refer to the extent to which an item is assumed to be known to the hearer, and can thus be referred to anaphorically and occur without a nuclear stress.

The distinction between 'given' and 'new' is used by Halliday, Chafe, and others to explain how it is that the intonation nucleus in a sentence is placed on the item with the greatest news value. Although some linguists (notably Chomsky and Halle 1968; Bresnan 1971) have attempted to account for nucleus placement in purely syntactic terms, it has been convincingly demonstrated by others (Bolinger 1972c; Crystal 1975) that nucleus placement cannot be accurately predicted unless semantic and situational factors are also taken into consideration.)

Traditionally, as noted above, members of the Prague School have treated givenness with theme, linking them more recently with the idea of communicative dynamism. Curiously Firbas, while insisting on the independence of given information from theme (1964:270), fails to separate the two notions consistently in his writings. (Chafe (1970:210-11) is another writer who combines the two notions. For him the given, or 'old' information is 'shared information', a kind of starting-point based on concepts 'in the air'. The approach adopted here is that in which, following Halliday (1967a, 1985, etc.), Trnka (1962) and Fries (1981), given and theme are clearly distinguished.)

5.3.1 *The realization of information structure*

(While most linguists are in agreement that information structure is realized – primarily at least – through intonation, there is some disagreement over how this occurs and precisely what intonational variables are involved. This section begins with an outline of Halliday's model, considers some points of disagreement between it and other models that have been proposed, and concludes with an outline of the position adopted in the present study.)

(Halliday regards discourse as being organized into a linear succession of message units which are realized phonologically as 'tone groups' (now more commonly known as 'tone units'). A tone unit has a phonetically specifiable contour organized around the 'tonic syllable' (that is, the syllable containing the greatest pitch movement and constituting (part of) the 'focused' word). The tonic syllable is the first syllable in the 'tonic (segment)'. Within the tonic segment the tonic syllable may be followed by further words which will continue the direction of pitch initiated in the tonic syllable.)

It is the internal organization of the tone unit which, argues Halliday, reflects the speaker's decision as to which constituent is focal or 'new' (that is, marked by the tonic) and which part is 'given' (that is, not within the domain of the tonic). In unmarked cases, the last lexical item in the tone unit is marked by the tonic as 'new' and the status of the rest of the information in the tone unit is not specified. In marked cases the tonic appears on some item other than the last lexical item, and the following information is assigned the status of 'given'. This difference is evident from the fact that, whereas an information unit with unmarked focus does not imply a specific *wh*-question, as in (13),

(13) # David bought the PAINTING # ('What happened?')

one with marked focus does imply a specific question, as in (14):

(14) # DAVID bought the painting # ('Who bought the painting?')

Halliday provides precise definitions of the terms 'given' and 'new'. For him they are interpretable not as 'previously mentioned' and 'not previously mentioned', but as 'information that is presented by the speaker as recoverable (Given) or not recoverable (New) to the listener' (Halliday 1985:277). (Halliday emphasizes in a number of places that it is the speaker who confers the status of givenness: 'It is his decision what to encode as given information and what to encode as new' (Halliday, unpub. MS §21.6). Subject though it is to the speaker's manipulation, the system of information is essentially listener-oriented: the message is organized in terms of what is and what is not news, with regard to its implications for the listener.)

Given status may be derived either from having been already mentioned, or from being physically present in the context of situation or generally known to all concerned in the culture. (See Section 5.3.4 below for detailed discussion of the categories of givenness associated with basic pseudo-clefts.) New status may be derived by being introduced into the discourse for the first time, or by being treated differently when introduced a second time. Halliday explains the difference as follows (1985:277):

'What is treated as recoverable may be so because it has been mentioned before; but that is not the only possibility. It may be something that is in the situation, like *I* and *you*; or in the air, so to speak; or something that is not around at all but that the speaker wants to present as Given for rhetorical purposes. The meaning is: this is not news. Likewise, what is treated as non-recoverable may be something that has not been mentioned; but it may be something unexpected, whether previously mentioned or not. The meaning is: attend to this; this is news. One form of 'newness' that is frequent in dialogue is contrastive emphasis.'

Thus Halliday treats contrastiveness as a kind of 'newness', and points out that an utterance like (15),

(15) # JOHN painted the shed yesterday #

can be both an answer to the question *Who painted the shed yesterday?* (where *John* is not contrastive), and to the question *Did MARY paint the shed yesterday?* (where *John* is contrastive). Halliday claims

that there are no reliable phonetic cues which will distinguish between these two functions.

While there is considerable agreement in the literature that intonation has a central role to play in the distribution of information across a text, there is not uniformity on a number of questions of detail. For instance there is little agreement as to what phonological variables expound the system of information. Whereas Halliday relates information distribution to tonicity, Brown *et al.* (1980) relate it to pitch. They recognize two pitch levels, each of which may be either 'boosted' or 'depressed'. 'New' lexical items are introduced at level 2 (the level for stressed syllables), contrasted or emphasized items are boosted to level 2+, and 'given' lexical items are depressed to level 2-. Level 1 is that assigned to unstressed syllables.

A different approach again is developed by Brazil *et al.* (1980), who associate information distribution with tone choice. They propose that the content of tone units containing a falling-rising tone (which they gloss 'referring') is 'given' (though the term 'given' is generally avoided because they wish to modify and expand the traditional concept) and the content of those containing a falling ('proclaiming') tone is 'new'. The rising and rising-falling tones are said to derive their meanings by contrast with the falling and falling-rising tones respectively. Brazil *et al.* are markedly at variance with the view of Halliday and most writers that the givenness of information within the tone unit is signalled by the absence rather than presence of phonological prominence, and that given information will always occupy a sub-part only of the tone unit.

Fronek (1983) and Firbas (1979) take issue with Halliday's claim that tone units in which nucleus placement fails to occur on the last lexical item are invariably informationally marked. They observe that the relative inflexibility of English word order means that informationally 'heavy' lexical items will not always be final in the tone unit. For instance a 'thematic verb' (that is, a verb of existence or appearance on the scene) may be relatively unnewsworthy, such that the unmarked instance is more likely to see the subject noun phrase (even if definite) carrying the nuclear tone. Thus (16a) would be a more natural response to the question *What has happened?* than (16b):

- (16) a. # your AUNT is coming #
b. # your aunt is COMING #

When the verb is highly predictable in terms of the subject, the nucleus falls most naturally on the subject, as in (17):

- (17) # the KETTLE's boiling #

Adverbials expressing concomitant information of various types may appear finally in unmarked tone units without carrying the nucleus. The following example is from Firbas (1979):

- (18) # you keep such a lot of RUBBISH in your bag #

Fronek's and Firbas's criticism of Halliday on these points reflects an inattentive reading of his work on intonation. Exceptions of the type noted here are commented on in Halliday (1967b:38), where it is observed that marked information is expounded by 'neutral tonicity' with 'certain high frequency collocations' and 'if the final lexical item is itself of very high frequency, at the grammatical end of lexis as it were'.

Another criticism of Halliday by Fronek (1983) which must be dismissed relates to Halliday's view that information distribution is subject to speaker manipulation, that it has a 'treated as' quality. Halliday argues that given information is not to be understood simply as previously mentioned material, but as material 'offered as recoverable anaphorically or situationally' (1967a:211). For Halliday, then, given information is an option which is 'not determined by the textual or situational environment' (1967a:211).

It is Fronek's view that recoverability (through anaphora and reference) is, or at least should be, an objective and quite automatic matter. However, as any linguist who has worked extensively with natural discourse will attest, information distribution is not simply predictable on the basis of what can be ascertained about the context. As Halliday suggests, the analyst must always be ready to allow that if the information structure is at odds with what might have been expected from the text, then the speaker will have had some good reason for wanting it to be so. Because the selection and presentation of information is determined by the speaker, the given-new distinction is available for exploitation: new information may be presented as if it were already shared (for instance for tactical reasons of politeness, or for one-upmanship), and given information may be presented as new (for instance because it is unexpected or needs to be attended to).

More deserving of serious attention is Fronek's (1983) criticism of Halliday's treatment of all elements carrying a marked nucleus as contrastive: 'An element marked by tonicity as contrastive may be either a lexical element that is not final or a final element that is not

lexical' (1967b:23). Fronek criticizes this formulation on two counts, claiming that:

- (1) it overgeneralizes, it assumes too many contrasts;
- (2) it fails to provide workable criteria for distinguishing between the types of contrasts vis-à-vis their real informational value.

In natural discourse, it does seem, there are often information units with marked focus which do not involve any noticeable kind of contrast: it is simply that the speaker wants to render an item especially prominent, or particularly emphatic.

(In this study I shall adopt a primarily Hallidayan approach and regard nucleus placement within the tone unit as the main determinant of information structure.) The distribution of information is not, of course, determined in each successive tone unit without reference to the linguistic and extralinguistic context. This fact is critical in informationally unmarked tone units, where (apart from the exceptional cases discussed above) the placement of the nucleus on the last lexical item merely marks the *culmination* of the new information. Some items within the tone unit will be marked as given by definiteness, deixis and various anaphoric devices. Marked information focus, where nucleus placement falls on other than the last lexical item, is regarded as carrying the meaning of 'contrastiveness', in the most general possible interpretation of the term. 'Contrastiveness' is understood to cover cases of special prominence or particular emphasis, even if there is no obvious sense of contrast involved.

It will be assumed that, as Halliday argues, information distribution, though listener-oriented, is subject to manipulation by speakers, who are at liberty to treat virtually any part of an utterance which they so choose as given. As Halliday (1970a:42) observes, nucleus placement 'can thus be used as a rhetorical device for implying that the hearer *should* have known or taken things for granted'.

One vexing question for models of information which relate information structure to intonational variables concerns the interpretation of written texts. Written language operates through a visual medium in which indicators of the writer's intended intonation typically do not appear in any systematic fashion, and yet writing somehow manages to encode information pertaining to what the writer assumes the reader does or does not know. One explanation might be that competent readers are able to derive such information via intonational patterns residing 'latently' in the text. Clues to such patterns will be provided by syntactic structure (for instance the correspondence, in

the absence of indications to the contrary, between clause and tone unit), and by punctuation.

An alternative explanation is that intonation plays a less significant role in the realization of information structure in writing than it does in speech. Such syntactic phenomena as definiteness and pro-form reduction will of necessity provide important clues to the reader, along with more general patterns of information flow within the text. This explanation does not require us to reject the possibility that a reader's sense of shared and unshared information derives, at least in part, from the latent intonation of a written text: it merely requires us to accord a more central role to non-intonational phenomena.

The suggestion that information status is determined in a less overt fashion in writing than speech derives some support from a study by Prince (1981) which suggests that readers are required to engage in more inferencing than are listeners in determining the givenness of discourse entities. Prince's findings are compatible with those derived from a comparison of the relationship between pseudo-clefts and givenness in speech and writing (see Section 7.4 below).

5.3.2 *The meaning of given and new information*

Having examined the manner of realization of given and new information, we need to consider what type of concepts the terms are used to cover. We have seen that Halliday approaches the question from the linguistic perspective of 'recoverability', where given information is that which the speaker presents as recoverable from prior linguistic context or from the situation, and new information is that which is presented as not recoverable. The concept of recoverability is employed by a number of linguists, including Kuno (1972, 1978, 1979), who defines given and new information as follows (1978: 282-3):

An element represents old, predictable information if it is recoverable from the preceding context; if it is not recoverable, it represents new, unpredictable information.

In addition to the notion of recoverability, three further concepts have been used in the definition of given and new information: 'predictability', 'consciousness', and 'shared knowledge'.

Information in the sense of predictability (as in the work of de Beaugrande and Dressler 1981) refers not to the content of communication, but rather to the aspect of unexpectedness associated with parts of a communication. At any moment the listener has certain

expectations about how the speaker is going to continue what s/he is saying. To the extent that the choice of an expression goes counter to (or rather perhaps is not predicted by) these expectations, this choice by the speaker is said to constitute new information.

Information is defined in terms of consciousness by Chafe (1974, 1976) as follows (1974:111-12):

What a speaker shares with his addressee must be part of what is in the speaker's consciousness at the time . . . The speaker knows he is not introducing material from his own consciousness into an empty vessel, but that his task is to introduce new things into a consciousness (the addressee's) which already has some content . . . The speaker must make assumptions as to what the addressee is conscious of, and transmit his own material accordingly.

For Chafe (1976:30) given information represents 'that knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance', and new information 'what the speaker assumes he is introducing into the addressee's consciousness by what he says'. Chafe's notion of givenness is the model for the notion used in Prince (1978) for the analysis of pseudo-clefts, except that it is extended to allow for any inferences so long as they can be deemed situationally appropriate.

A treatment of givenness in terms of 'shared knowledge' can be found in Haviland and Clark (1974) and Clark and Haviland (1977). Given information is defined as that which the speaker 'believes the listener already knows and accepts as true' and new information as that which the speaker 'believes the listener does not yet know' (1977:4). The hearer may know the information directly for having been explicitly told it, as in (19a), or indirectly via inferencing, as in (19b).

- (19) a. We got some beer out of the trunk. *The beer* was warm.
b. We got some picnic supplies out of the trunk. *The beer* was warm.

The four conceptions of givenness-newness outlined in this section, though different, are by no means unrelated. Thus, if a speaker assumes that the hearer can predict the occurrence of some particular item or items, then the speaker must assume that the hearer is able to recover these from the linguistic or non-linguistic context. This being so, the speaker must assume that it is appropriate that the hearer have some particular thing in his/her consciousness. If this is the case,

then the speaker must assume that the hearer 'knows', assumes, or can infer a particular thing.

5.3.3 Givenness and presupposition

The concepts of givenness and presupposition are closely related, both involving information which the speaker presents as known, or accessible, to the addressee. They are nevertheless distinguishable. Givenness refers to the informational status of individual linguistic items and constituents determined by the speaker's view of the linguistic and situational context, and is realized primarily by intonational means. By contrast presupposition involves a complete or potential proposition whose knownness or assumability is required for the success of the sentence. A presupposition is, in other words, a sort of 'given proposition'. Presuppositionhood is dependent upon a combination of syntactic and semantic factors, and is realized either as a dependent clause or as a constituent which may be reconstructed as a clause. Consider (20):

- (20) I have some trivia to offer you regarding the world's monarchs.
Let's start with France – the King of France is bald.

Here *the King of France* is clearly given, recoverable from *monarchs* and *France*. In itself the noun phrase is not a presupposition; what is here presupposed is the proposition 'There is a king of France.' Furthermore it is possible to have a proposition that is given, but not presupposed, if it is presented in anaphoric form following a statement that has asserted it. For example:

- (21) Tom offered Sue a sherry.
It can't be true.

Conversely, it is possible to conceive of an all-new (discourse-initial) utterance which presupposes a proposition. An acquaintance with a penchant for jokes and riddles might for instance begin a conversation with:

- (22) When will a net hold water?¹

Here nothing is given, but nevertheless a proposition ('A net will hold water at some time') is presupposed.

One aspect of the non-coextensiveness of givenness and presupposition that is particularly relevant to the analysis of basic pseudo-

clefts, is the common occurrence of informationally new items within a presupposition (a by-product of the unmarked coextensiveness of the clause with the tone unit). The 'freshness' of such items is nevertheless attenuated by the fact that they are syntactically 'backgrounded', presented as 'not-at-issue' (see further Section 6.1.1 below).

5.3.4 Taxonomies of information

Several taxonomies of given/new information have been proposed (Allerton 1978; Prince 1981). A feature that is common to these is the presentation of absolute givenness and newness not as merely binary terms, as in Halliday's model for example, but rather as poles on a scale (similar to the Prague School concept of a scale of 'communicative dynamism'). This section begins with an outline of two taxonomies suggested by Allerton and Prince. Then a classification is proposed.

Allerton's model incorporates four degrees of informativity, determined by the interaction of syntactic and intonational variables: 'new' (the speaker makes no assumption that the addressee knows which item s/he is talking about); 'semi-new' (the speaker requires the addressee to know which item s/he has in mind, but it can be recalled only by delving into the relatively distant past of the text or recesses of the mind (what Daht (1976:39-41) would describe as 'off-stage', and what Chafe (1974) would describe as not being in the hearer's consciousness); 'semi-given' (the addressee is required to be aware that the referent has been mentioned or made obvious in the fairly recent past (for Dahl 'onstage'; for Chafe, 'in the hearer's consciousness')); and 'given' (the addressee is referred to something in the immediate past context, and identification is taken for granted).

(In Prince's model, (which is applicable only to noun phrases occurring in discourse), there are seven values of 'assumed familiarity' which form a scale representing degrees of 'newness'. The seven values are grouped into three categories: 'new', 'inferable', and 'evoked' (where the latter two correspond to that which has generally been termed 'given' in the literature). 'New' is divided into 'brand new' (a newly-created entity, such as *a Mr Smith*), 'brand new, anchored' (a newly created entity linked to some other entity already in the 'discourse model', such as *a neighbour of mine*), and 'unused' (an entity known to the hearer, but not yet placed in the 'discourse model', such as *Mr Smith*). At the other end of the scale is the category 'evoked', referring to entities already in the 'discourse

model'. Such entities may either be 'textually evoked' (evoked earlier by the hearer, on textual grounds – that is, once 'new' or 'inferable', such as *the hicky devil* (referring to Mr Smith)) or 'situationally evoked' (by virtue of being situationally present, such as *me*). 'Inferable' entities are those the speaker assumes the hearer can infer from entities already in the 'discourse model', such as *the neighbour* (in the context of adjoining houses). A special subclass of inferables are the 'containing inferables' where 'what is inferred off is properly contained within the inferable NP itself' (p.236), such as *one of the neighbour's children*.)

Both Allerton and Prince unwisely tie their proposed categories to specific linguistic forms. As Brown and Yule (1983:187-8) have noted,

If we have to rely on linguistic forms alone to determine information status, it seems that the relevant status will not always be clearly marked and, indeed, if syntactic and intonational forms are both regarded as criterial for 'givenness', that these forms may supply contradictory information to the hearer.

(This problem does not arise in Halliday's model, where information status is taken to be determined by the speaker, rather than by the structure of the discourse.) Halliday (1967a:211) has emphasized that what would be focally marked as new is 'in the last resort what the speaker chooses to present as new, and predictions from the discourse have only a high probability of being fulfilled'.

The model presented below (see Table 5.1) has been designed to account for the categories of givenness that are associated with the relative clause of the basic pseudo-cleft construction. Givenness is not regarded as merely complementary to new information in a binary system; but neither is it, as in the Prince and Allerton models, regarded as merely occupying points along a scale. Rather, the basis for classification is the distinction between 'text-givenness' and 'situation-givenness' (Halliday and Hasan's (1976:33) 'endophora' and 'exophora' respectively), and a number of categories within each of these qualitatively different types are identified.)

(The present model has been influenced by aspects of all of the conceptions of givenness that have been described above, with Prince's (1978) study being a major stimulus. However, one difference between Prince's listing of categories and the taxonomy proposed here, is that the categories defined below are all theoretically motivated – formulated as values identified with non ad hoc textual and situational variables.)

Table 5.1 Categories of givenness for basic pseudo-clefts

	Co-textual		Contextual		
	Direct	Indirect			
Similar	Directly similar	Indirectly similar	FIELD	TENOR	MODE
Opposite	Directly opposite	Indirectly opposite	Material process	Mental process	Verbal process

As indicated in Table 5.1 the two major categories of 'sources' from which given information in the basic pseudo-cleft may be recovered – 'co-textual' and 'contextual' – together embrace seven subcategories. There are four co-textual categories, products of the intersection of two parameters (illustrated by the four examples in (23)–(26) below). The first ('direct'/'indirect') indicates the degree of effort, as it were, required of an addressee in order to retrieve the 'antecedent' for material presented as given.² Although any number of degrees of difficulty are theoretically possible (in other words a continuum of values rather than a binary opposition), the two values 'direct' and 'indirect' have proven sufficient to account for all corpus data in the present study. Whereas directly recoverable antecedents carry a high degree of co-textual predictability, and are thus easily assumable as present in the addressee's consciousness, indirectly recoverable antecedents require inferential 'bridge-building', and are distinctly less salient in the addressee's (assumed) consciousness. The second parameter which operates in the case of co-textual antecedents represents qualitative, rather than quantitative, aspects of the relationship between information in the relative clause and its antecedent. This parameter is more evidently binary in nature, divided between relationships based on similarity and those based on opposition. In (23) the given information in the *wh*-clause is directly similar to that in the antecedent indicated by dotted underlining; in (24) it is directly opposite; in (25) it is inferable that the discovery of something is comparable with lengthy investigation; and in (26) an opposition is to be inferred between a former rejection of women dentists and their current acceptance.

- (23) and I thought . I would jst DIE in this SÉT-UP # -- -- you KNOW # . I mean I'd RÉACHED the point # where I

thought ((well)) if they . if what would I do if they OFFERED me this thing # -- obviously what I'd do is TAKE it # (LL S.1.3.871).

- (24) # I am NÓT # for one moment ÁRGUING # that the king was PÓOR # -- or ever BECAME poor. # . all that I am ÁRGUING # . is the RÉLATIVE DECLINE # of the power of KÍNGSHIP # (LL S.12.2.566–8)
- (25) # well now Mr [naba] -- Mr Nabarro we will now ask you once MÓRE # -- the same question as we asked you BEFÓRE # -- [a:] . EXTENSIVE # and ÍNTENSIVE # enquiries have been MADE # into whether this . fear of this penalty . in fact deters people from MÚRDERING # -- and what we've DISCOVERED # -- STATÍSTICALLY # is that it DÓESN'T # (LL S.5.3.1013–15)
- (26) oh NÓ # . there were very FÉW # ((you know [am])) VERY few you had to PÁY # would never have made a LÍVING # because NÓBODY would have GÓNE to a woman dentist # -- it was because of the SHÓRTAGE # . why women are so accepted NOW # [?] IS # that because of the SHÓRTAGE of dentists {in ENGLAND #} # all the SCHÓOL dental officers # after the WAR # were -- because they were badly PÁID # were the WÓMEN # (LL S.1.13.81–7)

The model presented here is applicable to the description of informativity in basic pseudo-clefts, but does not work equally well for reversed pseudo-clefts or clefts. It is likely that the reason for this relates mainly to the difference in thematic structure between basic pseudo-clefts (relative clause = theme; highlighted element = rheme), and reversed pseudo-clefts and clefts (highlighted element = theme; relative clause = rheme).

The relative clause of reversed pseudo-clefts generally contains new information. However, the fact that it is expressed in a dependent rather than independent clause robs it of 'total freshness', as it were; that is, the information is presented as something which is not at-issue, something on which doubt or disagreement is not countenanced (see further Section 6.2 below).

Several writers have observed that there are communicative differences between information in the relative clause of basic pseudo-clefts and clefts (Prince 1978, Huddleston 1984, Gundel 1985). Because Prince's account is the most detailed, and has been the most influential, it is considered in detail in the following section.

5.3.5 *Givenness in clefts and pseudo-clefts*

In her analysis of the relative clause of basic pseudo-clefts, Prince finds a congruence of the notions of givenness and presupposition (p.887):

If we compare the logico-semantic notion of presupposition with the discourse notion of known/old/given information, we find that, in the case of the WH-cleft, they seem to match quite closely.

In fact Prince does not mention any case from her data of a pseudo-cleft in which the two notions do not correspond. However, as argued above in Section 5.3.3, givenness and presupposition are in fact separate notions (givenness is a discourse notion relating to the informational status of individual linguistic items and indicated by attenuated syntactic and phonological forms; presupposition is a logico-semantic notion indicated by sentence form and relating to assumptions which the speaker makes about what the hearer is likely to accept without challenge). Prince not only equates the two notions, but even refers to her practice in so doing as 'the usual custom' (p.886). However, one must question whether Prince's use of the term 'given' in this context does in fact correlate with the generally accepted practice of relating it to anaphorically or situationally recoverable material. For instance, in her account the capitalized *wh*-clause in the following example (= Prince 24c) is classifiable as 'given' (on the basis that it is relatable to preceding discourse material through a relationship of contrast, and is therefore assumable as being in the reader's consciousness).

- (27) The fact that . . . pre-eminence of some groups and regions over others shifted frequently is WELL KNOWN . . . WHAT IS LESS KNOWN, OR RATHER NOT ADMITTED BY SOME WHO PREFER NOT TO LOOK AT THE STARK PRESENCE OF REALITY, is the other fact that . . . (*Challenge*, p.25)

According to the approach adopted here, the *wh*-clause is certainly presupposed (that is, the writer requires the reader to accept the truth of the proposition it expresses), but at the same time it contains a mixture of given and new information (this observation following naturally, in an intonationally based account of givenness, from the fact that it would be intoned as several tone units). The fact that the new information lacks its normal salience is explained in the present

account in terms of the occurrence of the information in a clause which is presupposed, syntactically dependent, and which serves as the theme.

In Prince's discussion of cleft constructions there is no reference to the question of presupposition. In order to account for communicative differences between the relative clauses of pseudo-clefts and clefts, she proposes a distinction between two terms which are generally used interchangeably in the literature. 'Given information' is defined as 'Information which the cooperative speaker may assume is appropriately in the hearer's consciousness' (p.903), and claimed to be relevant to pseudo-clefts. 'Known information' is defined as 'Information which the speaker represents as being factual and as already known to certain persons (often not including the hearer)' (p.903), and claimed to be relevant to clefts. Prince contends that clefts mark the information in their relative clauses as 'known', but she interprets the concept slightly differently in the two major types of clefts. In 'stressed-focus *it*-clefts' (referred to in this book as 'unmarked clefts') the relative clause is said to represent information 'which is often, though not always, known from the context' (p.904). In 'informative-presupposition *it*-clefts' (which I shall refer to as 'marked clefts') the relative clause 'contains the "message" – but marked as a known fact, not as the speaker's assessment' (p.904).

In several subsequent articles, Prince has further explored the distinction drawn in Prince (1978) between 'given' and 'known'. In Prince (1979, 1981) three major types of givenness are proposed, of which givenness, in the 1978 sense, corresponds to 'saliency' ('The speaker assumes that the hearer has or could appropriately have some particular thing/entity/ . . . in his/her CONSCIOUSNESS at the time of hearing the utterance'), and knowuness corresponds to 'shared knowledge' ('The speaker assumes that the hearer "knows", assumes, or can infer a particular thing (but is not necessarily thinking about it)').³ In Prince (1985) these two concepts are re-labelled using the names of the scholars who are said to be most closely identified with them. The distinction between current attention focus (or, what is in the hearer's consciousness) and general knowledge (or, what is in the hearer's general knowledge store) is referred to by Prince as that between 'Chafe-givenness' and 'Clark-givenness'.

(An alternative to Prince's approach of manipulating the notion of givenness – adopted in this book – is to retain a constant and more narrowly-defined view of givenness, and attempt to explain the differences between pseudo-clefts and clefts in terms of the mapping of different discourse variables on to the given information in each case.)

In referring to a particular type of information which, in the case of marked clefts, 'contains the message', using a term ('known') that is generally interchangeable in the literature with 'given', Prince confuses two distinct concepts. This is not to suggest that the distinction between givenness and newness is a clearcut, necessarily binary, one. It is true that the freshness of the information in the relative clause of a marked cleft is attenuated by several factors: it is backgrounded through the dependency of the relative clause and as a result of the fact that it is logically presupposed, and imbued with a hedged, reportive flavour by the impersonal associations of the predicating *it* *be* structure. However, it would seem preferable to conclude that what we have in such cases is new information modified by various factors and in some such cases 'semi-new' information, rather than 'not new at all' (at least, to those excluding the addressee), which is implicit in Prince's use of the term 'known' for such information.

One of the reasons why Prince finds it necessary to extend the notion of givenness is that she fails, as she herself acknowledges (see Prince 1978:905), to incorporate the notion of theme into her account. The special communicative flavour of the relative clause of basic pseudo-clefts, in which the speaker appears to be making assumptions about notions that are, or could be, in the hearer's consciousness, results not from the expression of a special form of givenness but from the conflation of the functions of givenness (with presupposition) and syntactic dependency with that of theme, in the context of an intensive equative structure.

Aspects of Prince's account of the communicative differences between pseudo-clefts and clefts are challenged by Declerck (1988) and Gundel (1985), but their own accounts are open to a number of criticisms. Declerck questions Prince's claim that material in the relative clause of (basic) pseudo-clefts must always be given, on the grounds of their occasional use as discourse-openers. He concludes (p.215):

It goes without saying that the WH-clause of a cleft (of any type) that is used discourse-initially cannot present 'given' information (i.e. information that the hearer is assumed to be thinking about) and must therefore be informative.

Declerck's assumption here is surely ill-founded. There are contexts in which a speaker may quite cooperatively open a discourse with a basic pseudo-cleft, anticipating that the addressee will be thinking about material in the *wh*-clause. For instance, if I initiate a service encounter exchange with *What I'm looking for is a present suitable for*

a 6-year-old, it's likely that I do so expecting the shop assistant to have in mind that 'I'm looking for something.'

Declerck extends Prince's distinction between 'stressed focus' and 'informative presupposition' clefts to pseudo-clefts, arguing that the distinction is applicable to 'the larger class of structures to which both belong, viz. specificational structures' (p.216). However, while informative-presuppositionhood is undoubtedly a property of many reversed pseudo-clefts, it must be said that Declerck's basic pseudo-cleft examples are less than convincing. Consider (28) (= Declerck's (15)):

- (28) A I hear you've got a job at Johnson's. A nice place that is. I suppose you're happy now?
B Well I don't know. *What I'd really like to do is run a business of my own.* But I can't do that because I've no money of my own.

Much of the information in the thematic *what*-clause here (which is comparable in many ways to that in (14) above) is largely recoverable co-textually. It is no doubt the thematic status of the presupposed relative clause of basic pseudo-clefts that explains its lower tolerance of an informative presentation.

For the clefts and pseudo-clefts which he classifies as having an informative-presupposition, Declerck adopts Prince's 'new information represented as known' interpretation of material in the relative clause. One source of the knownness, claims Declerck, is the presupposition associated with the noun phrase expressing the variable (i.e. the relative clause).⁴ If this were so, it would be difficult to explain why the similarly definite variable in the 'stressed-focus' construction should carry such a different interpretation (i.e. 'given', or assumed to be in the hearer's consciousness at the time of speaking, rather than 'known'). In the analysis proposed below, 'not-at-issue' status, one of whose sources is presuppositional, is invariably associated with the relative clause of clefts and pseudo-clefts.

Declerck subdivides Prince's informative-presupposition category of clefts into two groups ('unaccented-anaphoric-focus', in which the highlighted element is old, and the relative clause is new but represented as old; and 'discontinuous', in which both the highlighted element and relative clause are new). These categories, along with the 'stressed-focus' type (which Declerck renames 'contrastive') do successfully account for the major types of clefts; a similar tripartite classification of clefts is presented in Section 5.3 below. However,

Declerck unwisely attempts to extend this classification to basic and reversed pseudo-clefts as well. He himself notes the 'frequent use of *it*-clefts where WH-clefts would not be appropriate' (p.242). In fact, not only can basic pseudo-clefts never have an accented anaphoric focus, as Declerck admits (p.224), but it is also questionable – as I have argued above – whether they can ever be 'discontinuous'. By contrast, reversed pseudo-clefts are very rarely 'contrastive' or 'discontinuous': corpus-examination indicates that they are generally in complementary distribution with basic pseudo-clefts.

Gundel's (1985) analysis invokes the notion of topic (a notion which differs from that of theme, as understood in the present study, in two major respects: unlike themes in English, topics are necessarily unstressed; and topics may be non-initial in the clause). In Gundel's view (basic) pseudo-clefts differ from clefts in that 'the expression which refers to the topic is in sentence initial, subject position in the former construction, but in sentence final position in the latter' (p.97). Sentence-final topics, claims Gundel, 'typically refer to some entity which is not only familiar to the addressee, but which is also "activated", i.e. an entity which the speaker's and addressee's attention is already focused on' (p.97). Gundel thus takes issue with Prince's account of the difference between *wh*-clefts and *it*-clefts, arguing that 'it is the *it*-cleft and not the *wh*-cleft whose relative clause must contain material that is already "in the addressee's consciousness" at the time of utterance' (p.98). As evidence, Gundel cites the following example (= Gundel (i) in footnote 18):

- (29) A How am I going to get this spot out of the rug?
 B What my mother always uses is vinegar.
 ?It's vinegar that my mother always uses.

As I have already demonstrated in the discussion of pseudo-clefts above, it is not necessary to have recourse to the (somewhat inaccessible) notion of the 'hearer's consciousness' in order to explain the discourse behaviour of cleft constructions. We can proceed a good distance equipped just with non-psycholinguistically-interpreted notions of theme and information. In Gundel's example, the pseudo-cleft constitutes a coherent response because its theme (*What my mother always uses*) is appropriately thematic in the exchange, providing a method of development in terms of the operation of removing spots. In informational terms, *what* and *uses* are recoverable; *my mother* and *always* are non-recoverable, and yet not entirely fresh. Because the relative clause of basic pseudo-clefts has, as I have

argued above, a strong backgrounding role deriving from its dependency as an embedded nominal clause, its expression of a presupposition, and its thematic status, speakers can (more so perhaps than in any other English construction), present information in the relative clause as 'assumable' even when such presentation is not textually or situationally warranted. Thus, although the information that Speaker B's mother 'always uses something' is not recoverable, it can be inferred that a mother is the sort of person one might regard as a source of knowledge and good practice in the event of problems whose solution requires a 'handy hint' (hence *my mother*). Furthermore it is inferable that repeated application of a particular method is likely to relate to its having proven successful in the past (hence *always*). The use of a pseudo-cleft by Speaker B imposes upon Speaker A an obligation to assume the 'backgroundness' of the information in the relative clause.

There are several sources of malcohesion with Speaker B's cleft. It is thematically aberrant: the text is not 'about' vinegar. The high communicative dynamism of *vinegar* creates the expectation that the content of the relative clause will be entirely given. But, as we have seen, it is not. Finally, the impersonal tone of the cleft sentence, deriving from the predicating structure, is not consonant with the typically personal tone of casual dialogue as represented in (29).

A more subtle form of explanation is needed to account for the fact that the reversed form of B's pseudo-cleft, *Vinegar is what my mother always uses*, would be more acceptable than the cleft. The two sentences are thematically similar and may be assumed to be informationally similar (assuming an unmarked reading). And yet they seem to differ on more than simply the question of impersonality. The distinction drawn by Halliday (1967a:236) between textual and ideational prominence (as discussed in Section 5.2.3) is helpful in explaining the nature of this difference. The theme of a cleft sentence is highlighted through predication and contains new information in the unmarked instance, and as a result carries textual prominence (here 'Vinegar and nothing else is the theme of this sentence'). The theme of a reversed pseudo-cleft is highlighted through exclusive equation with identified material in the relative clause, and as a result carries an ideational prominence (here 'Vinegar and nothing else is what my mother always uses').

I have questioned the plausibility of Gundel's claim that, *contra* Prince, 'Chafe-givenness' is associated with clefts rather than pseudo-clefts. (And I have already questioned Prince's claim that the reverse obtains: it follows that the 'Chafe-givenness-Clark-givenness' distinc-

tion is here regarded as incapable of explaining the communicative differences between pseudo-clefts and clefts.) Gundel argues that the different discourse-initial possibilities of pseudo-clefts and clefts are to be explained in terms of the association of the 'sentence final topic' of the cleft with an 'activated entity' (that is, one upon which the speaker's and addressee's attention is already focused, or 'Chafe-given' in Prince's terms). She says (pp.97-8):

In particular, we would expect that the *wh*-cleft can occur at the very beginning of a discourse when the addressee's attention can generally not be expected to be focused on the topic, but that the corresponding *it*-cleft will not be appropriate in such a context. This prediction is borne out by examples like the following:

- (29) (At the beginning of a lecture)
 (29a) What I would like to talk about today is conversational implicature.
 (29b) \$It's conversational implicature that I would like to talk about today.

(where \$ indicates inappropriateness in context, as distinct from ungrammaticality).

It is not necessary to resort to a consciousness interpretation to explain the facts here. A ready explanation is available if we divorce the notion of topic (or 'theme') from that of (given) information. The theme of Gundel's (29b), even though phonologically salient, is *conversational implicature*, and the inappropriateness derives from the presentation of so much news in the scene-setting, discourse-structuring segment of a discourse-opening sentence. On the other hand it is easy for an audience to accept, as an opening gambit, that a lecturer 'would like to talk about something', whether or not their attention is 'focused' on this proposition. (Likewise the first person pronoun theme of the corresponding non-cleft, *I would like to talk about conversational implicature today*, is easy to accept, given the frequency with which talk consists of messages concerned with the speaker himself or herself.) A further source of awkwardness with the cleft in Gundel's (29b) is the frequency with which the highlighted element in clefts, when stressed, carries a contrastive interpretation (that is, presupposes a set of non-selected candidates). In Gundel's (29b) there is no such set available to, or inferable by, the listener(s). (An appropriate context for the cleft would be provided if, say, the lecturer had walked into the room and, before commencing

his lecture, displayed an overhead projection which read 'Lecture Topics: (1) Reference (2) Presupposition (3) Conversational implicature (4) Theme (5) Information'.)

Further examples provided by Gundel can be used as evidence against her own argument. She quotes an example (= Gundel 33) provided by Prince (1978) (= Prince 41a) in which the relative clause contains new information (that is, is 'non-topical', in Gundel's analysis). What Gundel fails to mention is that this example is borrowed from a text in which it occurs initially.

- (29) a. It was just about 50 years ago that Henry Ford gave us the weekend.
 b. When Henry Ford gave us the weekend was just about 50 years ago.
 c. Just about 50 years ago Henry Ford gave us the weekend.
 d. Just about 50 years ago was when Henry Ford gave us the weekend.

Gundel's account does not explain the difference in acceptability between (29a) and (29b). Her posited 'topical' distinction between pseudo-clefts and clefts would presumably predict that (29a) should be unacceptable, and (29b) acceptable. The reverse, however, is the case. The source of the awkwardness of (29b) is similar to that of Gundel's (29b), as quoted above: namely, the inappropriate assignment of thematic status – and particularly thematic status in the opening sentence of the discourse – to the main news-bearing section of the sentence. The temporal reference, *just about 50 years ago*, is lower in communicative dynamism, and therefore more appropriately assigned thematic status. A greater challenge to explanation is presented by (29c) and (29d), which do not serve as felicitously as discourse-openers as (29a) despite the fact that the temporal adjunct is thematic, as in (29a).

Prince (1978:898), in her comparison of (29a) and (29c), comments:

Were the first sentence [= 29a] not clefted, i.e. *Fifty years ago, H.F. gave . . .*, it would seem as though the newspaper had just discovered (or were pretending to have discovered) the information in the *that*-clause; the *it*-cleft, in contrast, serves to mark it as a known fact, unknown only to the readership.

Prince is on the right track here, but for the wrong reasons. Undoubtedly the source of the difficulty with (29c) is the presentation of 'high CD' material in a discourse-opening sentence, and it may

well be that an associated implication is that the newspaper has just discovered such information (although this is surely not an inevitable implication). Furthermore, the information that 'Henry Ford gave us the weekend', when realized as a relative clause which serves as rheme in (29a), does indeed assume a certain flavour of 'knownness'. However, the 'knownness-feature' derives, I have argued, from a combination of factors (including syntactic dependency, logical presupposition, and perhaps the general impersonality of the cleft construction). Such factors attenuate the freshness of the news in the relative clause, enabling it to serve more readily as a discourse-opener. At the same time the theme, *about 50 years ago*, though clearly not given information, is less newsworthy than the proposition expressed in the rheme. Readers familiar with journalistic convention will recognize the practice of opening articles with one or more circumstantial adjuncts.

Huddleston (1984:464-5) identifies and discusses two 'main uses' of cleft sentences (presented schematically, for the sake of clarification, in Table 5.2). A feature of the first use (the 'unmarked' cleft, in my terms), is claimed to be that 'the superordinate clause selects freely for polarity' (p.464). This characteristic is undoubtedly associated with the typically contrastive role of the highlighted element, a role which entails sorting the items which are or are not relevant members of a set.

Table 5.2 Huddleston's two 'uses' of cleft constructions: informativity and 'at-issueness'

	<i>Superordinate clause</i>	<i>Relative clause</i>
Use 1 (<i>'unmarked'</i>)	New	Given
	+ At-issue	+ Not-at-issue
Use 2 (<i>'marked'</i>)	a Given ('often' - especially if anaphoric nominal) +	New
		+ Not-at-issue
	b New ('sometimes' - especially if adjunct) +	New
		+ Not-at-issue

In Huddleston's second type ('marked clefts', in our terms),

the superordinate clause will not normally be negative, though negation is possible under certain conditions when there is an easily derivable positive implication, as in *It wasn't until he died that she came to appreciate the importance of what he was trying to achieve*, which implies *It was when he died that she came to appreciate the importance of what he was trying to achieve*.

(p.465)

Huddleston argues that the information expressed in the relative clause of a marked cleft, though new, is presented as 'something not at issue, something on which doubt or disagreement is not countenanced' (p.465). Thus, in the examples given in the quotation above, it is not whether she has come to appreciate the importance of his efforts which is at issue, but merely *when*. According to Huddleston the highlighted element of marked clefts 'will often be anaphoric or its referent otherwise given, especially when the relative element is part of the nucleus, rather than an adjunct of time, place, or the like' (p.465).

Unfortunately Huddleston does not elaborate the theoretical status of his notion of being 'at-issue' or 'not-at-issue'. It seems to be in some way related to the notions of new and given information. However, the notions are not entirely equivalent. While given information is always 'not-at-issue', it would seem that new information may be either 'at-issue' (this presumably being so if other things are equal) or 'not-at-issue' (as for instance in the case of marked clefts, whose syntactic dependency 'overrides' the newness of information in the relative clause).

Huddleston recognizes that 'not-at-issue' status (which he associates with the relative clause of pseudo-clefts as well as clefts) must derive from sources other than merely syntactic dependency. He says: 'It is a slight oversimplification to relate this to the subordinate nature of the clause, for although subordinate clauses typically have this property, there are certain cases where they don't' (footnote, p.465). (I have identified one such source already, namely the presupposedness of the content of the relative clause in cleft constructions.)

Huddleston compares the 'non-given-but-not-at-issue' information that can occur in the relative clause of the basic pseudo-cleft and marked cleft, and asserts that the former is 'very much lower in communicative significance, in "communicative dynamism"' (p.466). Huddleston adopts a very narrow 'co-textual recoverability'

interpretation of givenness when he elaborates this distinction with examples as follows (p.466):

Suppose, for example, that after some energetic exercise I say *What I need now is a long cool drink*: that I need something is not given in the sense of having been mentioned or established, but it is very easy to accept and hence is presented, in the thematic organisation of the message, as subordinate to the identification of the thing needed. I would be unlikely to say *It's a long cool drink that I need* if it were not given that I need something.

The point is not so much that the information in the relative clause is invariably 'easy to accept': rather, it is so presented by the speaker. The speaker can afford to be almost cavalier about the actual informativity of the content of the relative clause, because the syntactic, semantic and logical form of the basic pseudo-cleft construction will do the job of backgrounding this information, even if it is non-recoverable. The listener is thus given the impression that it would be uncooperative of the speaker to treat this information as other than part of the background and, as we have seen, is accordingly obliged in many cases to build an inferential bridge to some contextual or contextual antecedent.

Huddleston's cleft example, *It's a long cool drink that I need*, would be unmarked, *ceteris paribus*. The highlighted element, an indefinite noun phrase with pre-head modification would be new, possibly contrastively so, in a typical context such as the invented exchange in (30).

- (30) A Game, set, and match. You lose again. Bill. And you look exhausted too. Come on, let's go for a swim in the pool.
B Well, I'm not so sure that a swim is what my body needs right now: *it's a long cool drink that I need*.

In (30) it would be possible to substitute the basic pseudo-cleft *what I need is a long cool drink* for the cleft. It is difficult to see why, following Huddleston, we should regard the relative clause of the pseudo-cleft as being less communicatively salient than that of the cleft here. Huddleston's primary justification for his claim is the difference in the linear sequence of elements that occurs in the two constructions: being thematic, the relative clause of the basic pseudo-cleft will be less communicatively salient; being rhematic, the relative clause of the cleft will be more communicatively salient. However, this undoubtedly valid claim must be set against two equally valid

claims which suggest that (if we restrict our attention to unmarked clefts) the relative clause of the pseudo-cleft is likely to be *more* communicatively salient than that of the cleft. Firstly, material in the relative clause of the pseudo-cleft is rarely directly retrievable from prior co-text: more often the listener is required to draw inferences. Secondly, the fact that material in the relative clause of the cleft is generally directly retrievable, and consequently is often ellipsed, suggests that it is comparatively undynamic.

To conclude, let me summarize the main differences between the accounts given by Prince and Huddleston of the communicative features of cleft sentences (ignoring Gundel's description insofar as it is based on notions developed by Prince). Prince attempts, in her description of the informational differences between unmarked and marked clefts, and between clefts and pseudo-clefts, to handle disparate communicative features in terms of a single variable. This variable, givenness, is elaborated through the addition of the notion of 'known' information. The concept of known information embodies a paradox. Such information is so-named because it is 'already known to certain persons' (Prince 1978:903), in which respect it may be regarded as a special case of givenness. And yet, in 'informative-presupposition *it*-clefts' (or, 'marked clefts') it is precisely this 'known' information which represents 'NEW information on the discourse level' (Prince 1978:898). Huddleston's account resolves this paradox. He posits two quite separate variables, which we might refer to as 'informativity' and 'at-issueness'. These would appear to be realized differently, the former mainly intonationally, the latter mainly syntactically. While newness will, *ceteris paribus*, be associated with 'at-issueness', and givenness with 'not-at-issueness', the special flavour of information in the relative clause of marked clefts (characterized by Prince as 'known'), derives from the marked combination of newness and 'not-at-issueness'.

The classification offered in Figure 5.1 draws partly on the insights of Prince and Huddleston but offers, ultimately, a fresh approach to the informational analysis of cleft constructions. Four categories of informativity, which is assumed to be a property associated with sub-clausal constituents, are posited. The four categories represent distinct feature combinations, and may thus be represented in the form of a systemic network.

Although there are feature differences between the four categories, they have in common the function of expressing degrees of informativity, and may thus also be interpreted as occupying points along a scale. It is at the end-points of the scale that intonational (and

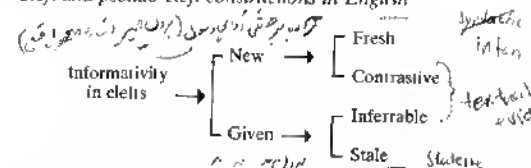


Figure 5.1 Informativity in clefts

often syntactic) clues are at their most reliable. Freshness is consistently associated with intonational salience, and often syntactic salience (the latter involving non-backgrounding through anaphora, deixis, pronominalization, and so forth), and staleness is consistently associated with intonational non-salience, and often syntactic non-salience. The absence of reliable intonational and syntactic markers for contrastiveness and inferability, the intermediate categories along the scale, is compensated for by an increased contribution from textual evidence.

Contrastiveness is regarded as a category of newness, but as less communicatively dynamic than fresh news on the grounds that an item so presented must have an element of givenness (that is, be recoverable textually or situationally). Contrastiveness is here understood to be realized either by marked information focus (in which case the primary sense is 'what follows is given', but this is generally further accompanied by a sense of contrast), or by nucleus choice (in particular, the choice of falling-rising rather than falling nucleus in an ordinary declarative clause, always used to signal a contrast). In the majority of cases the two means of realization will co-occur.

Inferability (which equates roughly with the category of the same name in Prince (1981)) is regarded, as in the discussion of pseudo-clefts, as a category of givenness based on a less-than-direct form of recovery of the 'antecedent'. With inferability there is sometimes a clash between intonational indicators (nucleus placement and choice) and textual recoverability. Typically, what happens is that there is a nucleus in a constituent that is evidently textually recoverable, and at the same time little evidence that information is being presented 'for noting'.

The relationship between the four categories may be defined, in terms of the concept of recoverability, in the following scalar fashion:

- Stale: recoverable directly
Inferable: recoverable indirectly (by inference)

Contrastive: recoverable, but to be freshly attended to
Fresh: not recoverable

There are sixteen theoretically possible combinations that arise from the association of the four categories of informativity, as defined above, with the highlighted element and relative clause of clefts. However, four of these may be discounted in view of the principle (of intonationally based theories of information) that all sentences must contain at least one item that is designated as news, either 'fresh' or 'for noting' (namely 'given + given', 'given + inferable', 'inferable + given', and 'inferable + inferable'). There were no instances of clefts representing any of these combinations in the corpus. The twelve remaining possibilities, and the frequencies with which they are represented in LL and LOB, are presented in Table 5.3.

Table 5.3 Informational classification of cleft constructions in LL + LOB

		Highlighted element	Relative clause	LL + LOB No. %	
UNMARKED	1a	New	Given	97	12.9%
	1b	Contrastive	Given	42	5.6%
	1c	New	Inferable	65	8.6%
	1d	Contrastive	Inferable	67	8.9%
	Total			271	36.0%
MARKED	2a	Given	New	221	29.4%
	2b	Given	Contrastive	15	2.0%
	2c	Inferable	New	22	2.9%
	2d	Inferable	Contrastive	2	0.3%
	Total			260	34.6%
REDUNDANT	3a	New	New	170	22.6%
	3b	New	Contrastive	3	0.4%
	3c	Contrastive	New	39	5.2%
	3d	Contrastive	Contrastive	9	1.2%
	Total			221	29.4%
Total				752	100%

The combinations (1a-d) are non-controversially classifiable as 'unmarked', and (2a-d) as 'marked' (see further Section 6.3.1 below). But how should (3a-d) be classified? If the criterion for defining a cleft as unmarked is the newness of the highlighted element, with no restriction upon the type of information that may occur in the relative clause, then (3a-d) are unmarked. If on the other hand the criterion for defining a cleft as marked is the newness of the relative clause, with no restriction upon the type of information that may occur in the highlighted element, then (3a-d) are marked.

It is the latter interpretation of the clefts represented by combinations (3a-d) that is adopted in both Prince (1978) and Huddleston (1984). The label that Prince uses for marked clefts, 'informative-presupposition *i*-clefts' reflects her definition of them in terms of the informativeness of the rheme, rather than the non-specified informativity of the theme. Prince does not comment upon the informational status of the highlighted (in her terms 'focused') material in 'informative-presupposition *i*-clefts', except indirectly in the observation that 'they generally have a short and anaphoric focus, which, in my data, is either a (subject) NP or an adverbial, generally of time, place, or reason - Kuno's "thematic scene-setting" adverbials (1975, 1978)' (1978:899) (compare p.904, 'the focus usually contains an anaphoric item'). Presumably the cases that we are presently discussing (as 'Type 3') are just those non-usual cases in which the highlighted element is non-anaphoric. For Huddleston, too, one of the two 'main uses' of clefts is definable in terms of the newness of information in the relative clause. Like Prince, Huddleston makes some brief and suggestive remarks about the informational status of the highlighted material in such cases: 'When the information in the relative clause is new, the highlighted element will often be anaphoric or its referent otherwise given, especially when the relative element is a part of the nucleus, rather than an adjunct of time, place, or the like' (1984:465). It follows that, in Huddleston's view, the highlighted element will *sometimes* be non-given, the most likely case being when the relative element is an adjunct of time, or the like.

The present data confirm the appropriateness of the (albeit sketchy) classifications offered by Prince and Huddleston. Type 3 clefts appear to have considerably more in common with marked clefts (that is, Type 2), than with unmarked clefts (that is, Type 1). In Type 3 clefts the main locus of news occurs in the relative clause, and the highlighted element, though clearly non-recoverable, is secondary in communicative significance. Typically, as noted by both Prince and Huddleston, the highlighted element is a circumstantial adjunct

(whose non-nuclearity, in an experiential sense, will tend to ensure its informational non-salience).

The alternative view (that which would lead to the description of Type 3 clefts as unmarked) is one in which the sole focus of attention is on the informational status of the highlighted element. (If such information is new, the cleft is unmarked; if not, the cleft is marked.) This view acknowledges the special prominence associated with the highlighted element in clefts through predication. Admittedly there is a certain attractiveness in this view; it leads to the classification of most clefts as unmarked (since all multi-tone-unit clefts in which the superordinate clause consists of one or more tone units will automatically be unmarked). However there is no necessary correlation between statistical frequency and unmarkedness (the latter notion being applied to a term in a system which is more 'basic', in terms of the number of features required for its description, than the other term(s) in that system). Furthermore, as already noted, the data used in the present study seem to indicate that Type 3 clefts have more in common informationally with marked than unmarked clefts.

I conclude this discussion by comparing twelve constructed examples representing respectively the twelve cleft-types presented in the classificatory scheme in Table 5.3.

- (31) a. In 1985 Mother Nature did not smile kindly on the countries of Central America. During the month of September Mexico City was hit by a catastrophe that reduced hundreds of buildings to rubble and killed thousands of people. On this occasion, it *was an earthquake of unrivalled intensity that struck the city.* ^{new}
- b. In 1985 Mexico had the misfortune to be afflicted not only by natural disaster, in the form of the massive September earthquake, but also by man-made disaster, with the continuing political upheavals. And which caused the greater loss of life? Thankfully, *it was not the political events that did so: the earthquake can lay claim to this dubious honour.* ^{given}
- c. During the month of September hundreds of buildings in Mexico City were reduced to rubble and thousands of people killed. *It was an earthquake of unrivalled intensity that was responsible for this catastrophe.*
- d. In 1985 Mexico had the misfortune to be afflicted not only by natural disaster in the form of the massive September earthquake, but also by man-made disaster, with the continuing political upheavals. Thankfully, it is not the politicians who

can lay claim to the distinction of having caused the greater loss of life. *It is the earthquake that merits this dubious honour.*

- (32) a. Another city to suffer the terrible vicissitudes of fortune in 1985 was Mexico City. *It was here that an earthquake of unrivalled intensity struck during the month of September.*
 b. People in most countries will probably remember 1985 as the year that millions of people starved to death in Ethiopia. However, Mexico City was the scene of a devastating earthquake, and it is *this that the people of Mexico will never forget.*
 c. In 1985 Mother Nature did not smile kindly on the cities of Central America, with Mexico's capital faring particularly poorly. *It was in Mexico City that an earthquake of unrivalled intensity struck during the month of September.*
 d. In 1985 the cities of Central America were struck by a series of natural disasters. *It was in Mexico City that the worst disaster occurred.*
- (33) a. *It was in mid-September that a natural disaster of unrivalled intensity struck the capital city of Mexico.* Thousands of people were killed and large sections of the city reduced to rubble.
 b. While people in most countries will probably remember 1985 as the year that millions starved to death in Ethiopia, *it is the devastation that occurred in the wake of the September earthquake that the people of Mexico will never forget.*
 c. Last year Mexico suffered a number of major catastrophes. In September the country's capital was struck by an earthquake that reduced hundreds of buildings to rubble and killed thousands of people. *It is this catastrophe that will undoubtedly go on record as one of the worst suffered by any country in 1985.*
 d. While people in most countries will probably remember 1985 as the year that millions of people starved to death in Ethiopia, *it is the catastrophe that occurred in their own country that the people of Mexico will never forget.*

In the cleft in (31a), representing Type 1, the highlighted element is an indefinite noun phrase containing new (fresh) information, and the relative clause is given because directly textually recoverable. In (31b) *the political events* is contrastive because newly identified as the entity that uniquely meets the description 'that which caused the

greater loss of life', in contradistinction to *the earthquake*. The relative clause, containing anaphoric items only, is clearly given. The cleft in (31c) differs from that in (31a) only in the informativity of its relative clause: that 'something was responsible for the aforementioned catastrophe' is inferentially recoverable via the simple deduction that if a catastrophe occurs, it will have been *caused* by something. The givenness is thus inferrable rather than explicit. In (31d), as in (31b), *the earthquake* is contrastively new, and the content of the relative clause is inferrable via the association between achieving distinction and meriting honours.

The cleft in (32a), representing Type 2, has as its highlighted element a demonstrative adjunct whose informational status is given because anaphoric; the relative clause introduces fresh information. The highlighted element of the cleft in (32b) is the given-because-anaphoric demonstrative *this*; in the relative clause *the people of Mexico* is contrastively new, the contrast being with *people in most countries*. In (32c) *in Mexico City* carries some informativeness, representing a type of 'semi-givenness': the addressee is expected to infer a relationship between *Mexico* and *Mexico City*. The relative clause, by contrast, introduces fresh information. In (32d) the addressee is expected to infer a link between *Central America* and *Mexico City*, while the focal item in the relative clause, *worst*, singles out the particular 'disaster' under discussion from a presupposed set.

In (33a) (understood to occur discourse-initially) the highlighted adjunct, *in mid-September*, represents fresh news; and so does the relative clause. As noted above, the relative clause, though the primary locus of news in the cleft, has an air of non-controversiality about it. It is the capacity of the Type 3 cleft to construe information in this way that explains, in large part, its popularity as a rhetorical device in certain genres (see further Section 7.4 below). In (33b) the highlighted element represents fresh news, while the content of the relative clause is informationally backgrounded except for *Mexico*, which is here contrasted with *most countries*. In (33c) the highlighted element, *this catastrophe*, is contrasted with *a number of major catastrophes*, and the content of its relative clause freshly new. In (33b) both the highlighted element and relative clause contain a contrastively new item (*own* and *Mexico* respectively).

In closing this section brief reference will be made to a suggestion which is developed at greater length in Chapter 7 below. If one takes into account, in discussing the informational structure of clefts and pseudo-clefts, the intersection of information and thematic structure, then one must conclude that the cleft construction is 'newness-

oriented' and the (basic) pseudo-cleft 'givenness-oriented'. With basic pseudo-clefts it has been found necessary to develop a classification rich in types of givenness in order to account for the construction's major function of delineating material in the thematic relative clause as recoverable. This classification has been found to be too delicate to be successfully applied to the description of clefts. By contrast a more delicate account of newness (subclassified into 'fresh newness' and 'contrastive newness') is called for in the description of clefts, but seems to be less necessary in the description of pseudo-clefts.

Reversed pseudo-clefts are a special case sharing features both with clefts (thematic highlighted element and rhematic relative clause) and pseudo-clefts (formally identical apart from the reversal of the highlighted element and relative clause). In reversed pseudo-clefts the relative clause typically contains the main point of newness in the sentence (even though, arguably, this is associated with the highlighted element in the unmarked instance). The typically given highlighted element, normally text anaphoric, is incapable of the degree of variety in types of 'antecedents' that is displayed by the relative clause of the basic pseudo-cleft.

6 Communicative meanings in the corpus

In this chapter the informational and thematic analysis of cleft and pseudo-cleft constructions that was developed in Chapter 5 is applied to the description of tokens from LL and LOB. The various communicative values that result from the interaction of given-new structure and theme-rheme structure are identified and, wherever relevant, figures derived from the corpus investigation are presented. It shall examine firstly basic pseudo-clefts, followed by reversed pseudo-clefts, and then clefts.

6.1 BASIC PSEUDO-CLEFTS

6.1.1 Intonation and information

I shall regard basic pseudo-clefts with a focal highlighted element as unmarked for two reasons. Firstly, the unmarked position for new information is at the end of the tone unit (an observation which will be relevant even if the pseudo-cleft consists of more than one tone unit – in such cases the highlighted element will still occur at the end of a tone unit, or perhaps occupy one or more tone units). Secondly, as with clefts, there is the unmarked correspondence between the 'identifier' and new information.

The figures for LL (as presented in Table 6.1) bear out Halliday's (1967a:226) claim, which is based on close textual observation, but not upon systematic corpus examination, that:

it regularly happens that an identifying clause is structured into two information units, the boundary between them falling, as usual in such cases, at the end of the theme: // the one who painted the shed last week // was John //

(Of the 272 basic pseudo-clefts, 249 (91.5%) consist of two or more tone units. In all but a handful of cases the theme is cointensive with the first tone unit. Thus in terms of theme and information structure (1) and (2) below are typical, (3) and (4) highly untypical.)

- (1) # what you WANTED was TWO THINGS # (LL S.1.4.245-6)
 (2) # - it's just that - - - what they KNOW about # is experimental RESEARCH # (LL S.2.4.727-8)
 (3) C what they had was a SURGEON # - talking to his - , the younger members of his staff as they went round the WARD #
 A (I DUNNŌ #))
 B I SAY # LISTEN #
 >C and talking to the PATIENTS # (LL S.2.5.94-5.99)
 (4) # what they found WAS # that it didn't make a damn bit of DIFFERENCE # - in FACT # - [a] . in certain CASES # in the matter of . people - committing . crimes AGAIN # (LL S.5.3.312-16)

Table 6.1 Nucleus placement in pseudo-clefts in LL

	Pseudo-cleft realized as single tone unit			Pseudo-cleft realized as two or more tone units		
	Nucleus located in theme	Nucleus located in rheme	Total	Nucleus located in theme ¹	Nucleus located in rheme ²	Total
Basic pseudo-clefts	3	20	23	224	25	249
%	1.1%	7.4%	8.5%	82.4%	9.2%	91.5%
Reversed pseudo-clefts	13	276	289	10	52	62
%	3.7%	78.6%	82.3%	2.8%	14.8%	17.7%

Notes

¹ Pseudo-clefts listed in this column have a nucleus in the theme, and one or more in the rheme (depending on the number of tone units).

² Pseudo-clefts listed in this column have no nucleus in the theme, and one or more in the rheme.

Of the 23 basic pseudo-clefts realized as a single tone unit the majority (20, or 87.0 per cent) have the nucleus located in the rheme/highlighted element. Thus (5) represents a typical example of such a pseudo-cleft.

- (5) # and all I'd do was get UPSET # (LL S.2.12.305)

However, as Table 6.1 indicates, in the majority of basic pseudo-clefts the relative clause/theme contains (at least some) new information. Nevertheless there are several factors which work together to attenuate the newness of the item or items so marked intonationally. As an embedded clause, the relative clause of basic pseudo-clefts represents information that is syntactically backgrounded, and thus imbued with a non-controversial flavour. Furthermore this clause represents a presupposition, a proposition assumed to be true by the speaker. Finally, the information focus in question falls within territory which, being explicitly marked as thematic by the pseudo-cleft construction, has its prominence (be it 'freshly new' or 'contrastive') attenuated by reason of the regular association of theme with given information. (The fact that the theme of basic pseudo-clefts is structurally unmarked, representing as it does the subject function, diminishes the possibility of its conveying the strongly new flavour which is regularly associated with marked themes in English.) Thus the focally new item or items are generally prominent only to such a mild degree that they do not detract from the backgrounded character of the relative clause as a whole. The relative clause generates a strong feeling in the addressee that s/he is being expected to accept, or infer the relevance of, the proposition presupposed by that clause, while at the same time being expected to pay particular attention to a section of that proposition which the speaker highlights prosodically.

The classification of information for basic pseudo-clefts that was presented in Section 5.3.4 above, although it does not accord a special place to intonation, is thus not incompatible with intonationally based accounts of information distribution. Several examples may clarify the interaction of presupposition/syntactic backgrounding and intonationally-realized information structure. In (6) the only element of newness in the theme is the indication of affirmation (*really*): all else is directly recoverable from prior linguistic context. The intonational prominence attributed to the adjunct does not detract from the status of material in the relative clause as presupposed. Expressed in terms of the concept of communicative

dynamism, we might say that the contribution of information in the theme of the pseudo-cleft is far outweighed by that in its rheme.)

- (6) (. . .) # – the only thing I ever VARY # . you CAN vary # is really [?] well you can vary ANYTHING # but the only thing I'm [?] – the thing that you REALLY vary # is [ði] – HOPS # (LL S.1.7.253–4)

In the vast majority of basic pseudo-clefts in the corpus whose theme contains a nucleus, the nucleus falls upon an item that is retrievable from the linguistic or situational context, either directly or inferentially. It follows that the 'newness' associated with such items will be of the contrastive rather than fresh variety. (Here 'contrastive' is used in a more general sense than by Halliday, who reserves it, in the context of discussing information structure, for items in the tone unit other than the final lexical item which carry the nucleus.)¹ The meaning of the information focus is here simply 'pay attention to this item, because I believe that it requires special emphasis'. The reasons for this 'flag waving' may be anaphoric (contrasting with a previous assertion), or cataphoric (contrasting with what one might expect).

In (7) *appeals* is synonymous with *like*, yet carries a falling-rising nucleus. The function of the nucleus is partly to sharpen the focus of A's response to B's question ('and now I'll tell you why I like Latin . . .'), and perhaps also to reinforce the positiveness of the response (by suggesting a contrast between 'strong appeal' and 'mild appeal': 'what really appeals . . .').²

- (7) B do you LIKE Latin #
A yes I DO quite # [a?] what APPEALS to me about it is that [əm] # . you have a great long string of ENGLISH # that you . make into about . THREE # – . well though out WORDS # . I like the PRECISION of it # (LL S.3.1.666–9)

It is because focal items in the relative clause of basic pseudo-clefts are typically 'new' in this mildly contrastive sense that there is no sense of incongruity with their participation in a proposition that is semantically presupposed and backgrounded through its syntactic realization as a dependent clause. (There is no contradiction between assuming the truth of a proposition and drawing particular attention to a section of it.) In fact it is the 'tension' that results from the conflation of apparently incongruous elements representing different linguistic systems which, in conjunction with other mappings, gives

rise to the unique communicative meaning generated by the basic pseudo-cleft construction. As so often happens in grammar, new and subtle meanings are being generated at the frontiers of the semogenic process.

6.1.2 Informational classification of basic pseudo-clefts

In this section the seven categories comprising the model proposed for the informational description of basic pseudo-clefts in Section 5.3.4 will be discussed in greater detail and exemplified by data taken from LL and LOB. The relevant frequencies for the two corpora are presented in Table 6.2.

Table 6.2 Pseudo-clefts and givenness: frequencies in LL and LOB¹

Corpus	Co-text				Context		
	Directly similar	Directly opposite	Indirectly similar	Indirectly opposite	Field	Tenor	Mode
LL	1.2(53)	0.6(26)	0.9(41)	0.4(18)	1.4(62)	1.1(46)	0.6(26)
	3.2(138)				3.1(134)		
LOB	0.1(10)	0.3(27)	0.6(59)	0.3(28)	0.2(22)	0.3(33)	0.1(13)
	1.2(124)				0.7(68)		
Total	0.4(63)	0.4(53)	0.7(100)	0.3(46)	0.6(84)	0.6(79)	0.3(39)
	1.8(262)				1.4(202)		

¹ In Table 6.2, and subsequent tables where indicated, frequencies given represent the number of tokens per 10,000 words (in order to compensate for the unequal size of the two corpora: 435,000 words in LL and 1,000,000 words in LOB), while bracketed figures represent raw frequencies.

(i) Directly similar

Here the source for the information backgrounded in the relative clause is represented explicitly, or more or less so, in the preceding linguistic context. Directly similar antecedents are almost four times as common in the spoken as in the written data, a fact which may be attributed to the greater tolerance of redundancy (in this case, lexical repetition) associated with the former. Writers can well afford to trim instances of redundancy from their language as they move from first draft (notes, pre-writing conceptions, and so forth) to final product,

given the absence of temporal restriction that typically characterizes the event of its decoding. However, in the case of speech the listener has only as much time to decode the discourse as the speaker has to encode it (or perhaps less, if the speaker has had the opportunity to premeditate or prepare).

Consider (8):

- (8) D AND [ə:] # . WHÁT was it # that he really wanted to DÔ # -
 B go on the little CARS # (- laughs)
 c [m] #
 D he wanted to sit in one of those little RAILWAYS # that go
 round and round in a figure EIGHT # - and the thing that he
 wanted to sit in was a POLICE car # (LL S.1.12.1420)

Here Speaker D recollects a visit to a safari park during which her nephew revealed a preference for riding on the railway over viewing the animals. In the first tone unit of D's final turn in the extract she picks up the earlier observation about the child's desire to ride on the railway (an observation initiated by herself, and completed in anticipation by B). The proposition that 'he wanted to sit in something' is then explicitly reiterated, in phonologically non-salient form, in the theme of the pseudo-cleft, in which D more precisely identifies the type of car in which her nephew wished to ride.

The value of a model of given/new information based on notions of recoverability from context becomes evident in the analysis of written examples, in which the clues of intonation are not present. In (9), *the thing that Serena discovered* has an explicit antecedent in the preceding paragraph (and in the context of the chronological sequence of events being described, *next* is also anaphoric).

- (9) Fortunately, however, the fashion for Victorian architecture which Mr John Betjeman had started several decades before had caught on at last and therefore saved the situation for the affluent middle-class, who now had plenty of lovely-ugly to be coldly elegant in.

All this Serena discovered, and more, but in stages. For the first thing she did was to make an offer on a small pink terraced cottage, two beds, two inter-comm. rec., mod. k. and b., sep. W.C. small back yard, newly dec., near shops and tube in up-and-coming Camden Town Village, £6,000 Freehold.

The next thing that Serena discovered was that she could not afford to buy a house at all. (LOB RO3.67-8)

(ii) *Indirectly similar*

With this category the addressee is called upon to build an inferential 'bridge' to an antecedent of similar meaning (but not of similar form, as is often the case with directly similar antecedents). In terms of scalar models of given-new information, such as those presented in Section 5.3.4 above, this category falls somewhere between the poles of the continuum, though certainly closer to the 'given' end. In most cases the inference is of a type that presents few difficulties for a cooperative addressee. For instance in (10) it is readily inferable from the fact that the speaker has a number of queries of different kinds to which he would very much like the answer that 'he wants to know something'. The familiar tension between intonational prominence and textual recoverability - as described above - is in evidence in this example (*know* carries a nucleus, yet its content is readily inferable). The meaning thereby generated is paraphraseable as 'Pay attention now' ('... for I've come to the first of the queries which I'd like answered').

- (10) # this is SÍMON Lessin SPÉAKING # good MÓRNING # TÔ
 you # . I have . a number of QUÉRIES of different KÍNDs # .
 to which I would very much like the ÁNSWER # (. .) # what
 I want to KNOW # . ÍS # . is the SPÉLLING # - on the TÍTLE
 # . L U C I L Í # . ar is it L U Ć # . I L DÓUBLE I #
 (LL S.9.3.464-70)

In (11) the process of inference which the addressee is called upon to make is clearly represented. Speaker D is required to accept Speaker A's deduction that, since D will not receive total reimbursement for his conference in Argentina, he will have to 'produce something'. The actual amount which A deduces that D will need to produce is then presented as news in the rheme.

- (11) D # [ə:] . WÈLL # . I was {GÓING to} GÔ # to a {CÓN-
 FERÉNCÉ} in ARGENTÍNA # (. .) # . but I can't get [ə]
 any FINANCES for it # - [ə:m] . EuroLang will pay
 THIRTY per cent # and the Foreign Survey - [ra:] rather
 late in the day said they hadn't got the FUNDS # . I SÁY
 THÁT # because Hilda is definitely AT(TÉNDING)
 THÁT conference # . and we HAD been . planning #
 maybe to [ə:m] - [ə] stop off [ə:m]
 A and is SHÉ # FINÁNCED to it #
 D yes yes she ÍS #

A YĒAH # . so *what* you've got to PRODŪCE # . is . in fact
thirty-five per CENT # of the joint EXPENSES # (LL
S.9.2,847-9)

In (12) the assumability of the proposition that 'something gives rise to optimism' is established by the semantic similarity between *outlook will be brighter* and *optimism*, in conjunction with the bridge that the reader must construct to the effect that, if optimism exists, then there is something that *gives rise* to it.

- (12) If however, in addition to her new good-neighbour gesture, Germany takes a really big share in giving aid to underdeveloped nations, the world outlook will be brighter.

What gives rise to optimism is the sign that Germany and the other leading Western nations are at long last moving towards a solution of currency problems by co-operation. (LOB BO1,87-9)

(iii) *Directly opposite*

Directly opposite antecedents differ from directly similar antecedents only in respect of a component of contrast. The contrast may involve (individually, or in some cases in combination) polarity, antonymy, degree, time, or the like. It typically involves marked information focus (such that, for example, the set of possible 'candidates' is restricted to two ('x' and 'not x')). However, there also occur cases where an information focus may be co-textually contrastive, though not informationally marked, and even cases of non-focal material expressing a contrast. In (13) there is a direct contextual contrast (between *would not trust the German nation* and *would trust them*), but not contrastive focus.

- (13) # . I PERSONALLY # would not trust . the German NATION # . as I've known it during my LIFETIME # . with a PÖPGUN # let alone with [ði:] REAL machinery # of WÄR # - *what I would trust them WITH # . is the machinery of PĒACE #* (LL S.5.1,689-90)

When the material that is being presented as 'retrievable through contrast' is focal, there arises the familiar tension between the sense of retrievability and the 'attend to this' sense of focus-marking. In (14) the operator *am* carries marked focus, so that the following lexical item, *convinced*, which is clearly co-textually recoverable, is

signalled phonologically as given information. The sense of oppositeness is reinforced by the choice of falling-rising nucleus which here expresses, as it regularly does, a contrastive meaning.

- (14) now this is a VĒRY complicated BUSINESS # . of {COURSE} it MÉANS that # a thing like a COM{PÜTER} is brought IN # - I'm FÄR from CONVINCED that the whole operation can be done in one great JAMBORÉE # . but *what I AM convinced is # that we shall never do anything unless departments COÖPERATE #* (LL S.3.4,548-9)

In the next example, from the written corpus, it is once again a negative-affirmative contrast that provides the link between material in the relative clause and its antecedent. It is probably safe to assume that most readers would intone a falling-rising nucleus on the focal item (arguably *not*).

- (15) The Labour Party, as we know, is in decline. *What we may not know is that the Tory Party is also in decline.* (LOB RO1,145-6)

(iv) *Indirectly opposite*

Indirectly opposite antecedents are related to material in the relative clause of the pseudo-cleft by means of an opposition that the addressee is required to infer. As with the previous category, the opposition may be supported by contrastive focus, but it may also be associated with unmarked focus or non-focal material. For example in (16) *quiet*, which carries unmarked focus, contrasts with *a terrible mess* earlier (and, more generally, with references to trouble in Ireland in the preceding discussion); *before*, which is represented intonationally as given information, not only picks up B's earlier reference to 'always (been in a mess)', but also contrasts with 'since (nineteen sixty-eight)' in A's first turn in the extract.

- (16) B but SÖRRY # ((4to5 sylls)) what would you DÖ about it # I mean would you - you [s] I mean you say that the border is IMPOSSIBLE # but how would you REMOVE it # - would would you . would you make Ireland into one.
A WĒLL # it's all in a terrible mess NÖW # I mean I don't know what I'd do . NÖW # since nineteen sixty - . EIGHT # - but
B (([m] actually that's not)) no that's not that's not the PÖINT # ((2to3 sylls)) it's ALWAYS been in a mess # in a SENSE #

f [m]

A ((2to3 sylls I don't think))

B you KNOW # I mean there's been no easy TIME # the reason it was QUIET before # nineteen SIXTY-EIGHT # was because , you can ARGUE # is because , the British - didn't , didn't stir up the Northern [aɪər ə'zʌm] the Ulster PROTESTANTS # (LL S.2.8,431-4)

Whereas the predominant type of contrast represented with directly opposite antecedents is that of polarity, indirectly opposite antecedents involve a number of different types of contrast. In (17) the contrast is based on the antonymous relationship between 'forget' and 'bring back' which enables the pseudo-cleft to thematize material that is both, in one sense, 'in the air', and that can legitimately be presented as deserving of special attention through focalization.

- (17) After this I visited Mr Smith very seldom, and if my memory serves me correctly, the only additional secrets I handed over were plans for a new-style cardigan, a patent cycle hmb-cap, a beer-engine and some air-line cutlery. Our mesalliance slowly collapsed and until last week I had almost succeeded in forgetting all about it.

What brought it back were the recorded impressions made by BBC reporters of their May Day visit to Moscow.
(LOB R04,82-3)

In (18) the contrast is based on degree (*has long been known* as opposed to *is much less clear*).

- (18) William Garnett and Robert Morant (who drafted the Bill of 1902 and was rapidly promoted to become first permanent secretary of the new Board of Education) have shared the same friendly biographer: and it has long been known that these two men were at the centre of the network of political activity within which steps were taken to ensure that the ball went on rolling. *What is much less clear is how far Sidney Webb was responsible for the direction taken by affairs in 1901 and 1902, and how closely he worked with Morant, with whom he and Beatrice were of course later to quarrel over social insurance.*
(LOB J39,96-9)

(v) Field

It was argued in Section 5.3.4 above that this category is represented by two subtypes. In one, the relative clause of the pseudo-cleft

contains a form of pro-verb *do*, and in the other a form of pro-verb *happen*. Field antecedents are generally not retrievable from anything in the preceding co-text; rather, they are reconstructable in terms of features of the non-linguistic context, and our knowledge of the world. Field antecedents point to the existence of a pragmatic principle that our experience of the world consists of a series of 'doings' and 'happenings'. The important thematic function of pseudo-clefts with relative clause containing pro-verb *do* or *happen* is discussed in Section 6.1.3 below.

Do and *happen* are the only members of the pro-verb class. Of them Halliday and Hasan (1976:125) say:

These stand for any unidentified or unspecified process, *do* for actions and *happen* for events (or for actions encoded receptively, in some kind of passive form). Their occurrence does not necessarily involve an anaphoric or cataphoric reference.

As pro-verbs, the *do* and *happen* that occur in this category of pseudo-clefts have features distinct from those of other uses of these verbs. Consider the various uses of *do* exemplified in (19).

- (19) a. What's Tom doing in Darwin?
b. What he's doing is working on a research project.
c. He's doing a project on lexical variation.
d. The heat won't do him any good.
e. He can always stand it better than I could do.

The pro-verb use of *do* is exemplified in (19a) and (19b), where the process referred to by *do* is unspecified by virtue of its being, respectively, the 'object of enquiry' of a *wh*-question, and the 'identified' element in an equative construction. The role of pro-verbs in pseudo-clefts, as in (19b), is discussed further below.

The pro-verb use of *do* is to be distinguished from its use as a lexical verb, a general verb, and a substitute. Lexical *do*, as exemplified in (19c), is an ordinary verb. It is not cohesive, except in the special context of lexical cohesion. General *do*, as exemplified in (19d), is a lexical item with generalized meaning. It is a member of a small class which also includes *make*, *have*, and *take*. Like lexical *do*, general *do* is only cohesive in the context of lexical cohesion. Substitute *do*, as exemplified in (19e), manifests a relation of presupposition at the lexicogrammatical level. (In (19c) *do* presupposes *stand* in the previous clause.) Substitute *do* is almost always anaphoric. (In (19e) the two clauses are structurally related, so that the presence of *do* is not essential as a cohesive factor; but it frequently

substitutes – at least in British English – for an element in a preceding sentence, serving in such cases as a primary source of cohesion.)

It is necessary to distinguish two subclasses within the 'field' category: that in which *happen* rhematizes the entire clause representing the ignate non-cleft, and that in which pro-verb *do* or *happen* is accompanied by one or more participants in the relative clause. In the first case *happen* may denote a virtually unrestricted range of 'goings-on' reflecting our experience of the world. Thus the 'goings-on' may be actions and activities (*what happened was that Tom flew to Darwin*), psychological processes (*What happened was that Tom liked the city sufficiently to want to extend his stay*), behaviours (*What happened was that Tom fainted in the heat*), verbalizations (*What happened was that Tom asked for an extension*), and even 'beings' (*What happened was that Tom was the only one to apply for Darwin*).

In the second case one or more of the participants in the process is included in the relative clause of the pseudo-cleft, the range of process-types is restricted to the 'material' and 'behavioural' types (that is, actions and activities), and the nature of the participant(s) involved determines the choice of pro-verb *do* and *happen* as follows:

- (i) For actors (and 'behavers') alone *do* is used, whether the action is voluntary or involuntary (*What Tom did was fly to Darwin*; *What Tom did was faint in the heat*). In addition a goal may be specified in the relative clause (*What Tom did to his winter clothing was burn it*). Psychological processes and 'beings' are not permissible (**What Tom did was like the city*; **What Tom did was be the only one to apply for Darwin*).
- (ii) For goals alone *happen* is used (*What happened to Tom's winter clothing was that he burned it*). Sometimes 'behavers' can occur with *happen*, as well as *do* (*What happened to Tom was that he fainted in the heat*). Psychological processes and 'beings' are of dubious acceptability (?*What happened to Tom was that he liked the city*; ?*What happened to Tom was that he was the only one to apply for Darwin*).

As with the textual categories, so with the contextual categories there is commonly a tension set up between the intonationally realized focus of information and the backgrounded nature of material in the relative clause. Consider (20):

- (20) (...) # . I HĒARD # . the HEADMĀSTER # [?] and the history MĀSTER # discussing . ESSAYS # ((5to6 sylls)) they

were quite STUNNED # by[di:] . MATŪRITY # and PROFUNDITY # and , the {DEPTH} of his VIEWS # – and I listened to this DISCUSSION # with a – sort of with . ((a little)) {CYNICAL} SNIGGER # INSIDE me # . I went to this OTHER person # . some years LATER # . and COMMENTED on this this this # and said that I presumed that what he'd DŌNE # was to very CAREFULLY # note down [?] every VIEW # he'd heard THEM express # . and serve it up in his own WORDS as {THEIR views #} # (LL S.1.6.321–5)

Here the newness conveyed by the falling–rising nucleus on *done* is contrastive in flavour. It is hardly fresh news that in submitting an impressive essay the student in question had 'done' something. However, the commonly contrastive meaning associated with the falling–rising nucleus is being exploited here to establish an opposition between the apparent and the actual (the illusion, in this case, being the impression given by the student's essay of a mature intellect). This element of newness does not prevent us from analysing the proposition encoded by the relative clause as 'not-at-issue' (because backgrounded and presupposed). In any event, it is clear that the 'message' conveyed by the pseudo-cleft is located in its rheme. The pseudo-cleft construction enables the speaker to establish a clear opposition between a past event and the later revelation of what had actually happened. This is effected by grouping sentence elements in such a way that the process is part of the highlighted element, and referring to it cataphorically by a form of *do* in the relative clause. It then becomes possible to exploit the possibility of a polar contrast inherent in the *do*. This sense of contrast is further reinforced by the exclusive identity feature of the construction, whose 'identifier' introduces a candidate to satisfy the description provided by the 'identified element' which is set against other possible candidates in the discourse.

With field antecedence of the *do*-type it is often evident that the addressee is intended to retrieve an antecedent from the context of situation in which the discourse is taking place. The example in (21) is from Text S.2.10 of LL, in which the speakers, having imbibed several glasses of champagne, attempt the feat of eating cherries without using their hands. As they do so they make references to the particular methods that they are employing, using basic pseudo-clefts. In (21) the action represented by *do* is to be retrieved situationally, so the nucleus-placement signals a contrast ('this is the way to

do it, not that way'). The sense of contrast is reinforced by the presence of *only* as premodifier to *way*, and the choice of falling-rising nucleus.

- (21) d ((I'll try again – like that)) (. sniffs) you put the eud in your mouth
 A ((there's)) {NÖ} WÄY of doing it #
 C ((no ÖUT #))
 B *the only way you can DÖ ((ü)) # is by sucking your tongue out to BEGİN with # and getting your tongue DRY # – and then LICKING it up with your TÖNGUE #*
 (LL S.2.10,1301–4)

Field antecedence of the *happen*-type is claimed by Prince (1978:893) to be consistent with the existence of 'a pragmatic principle that says that events keep occurring – and that in our culture, at least, they are our proper and constant concern'. In (22) the relative clause, as well as presenting the occurrence of some past event as a legitimate concern of the reader, derives a contrastive sense from the presence of *really* ('I am about to reveal the historical facts to you, in contradistinction to the fictitious legend which I have just described').

- (22) Hamelin is mainly familiar to us through the legend of the Rattenfänger (Rat-catcher), related in Browning's poem. The event is celebrated each Sunday in summer when the story is re-enacted by a piper and boys, the latter disguised as mice. Unfortunately, modern research tends to discredit the legend, claiming that *what really happened was a visit from a labour agent who attracted many local young men away to Bohemia, with the promise of good wages.* (LOB F35.57–9)

In the case of (23) the antecedent to 'something has happened', though not explicitly retrievable from the preceding context, is nevertheless quite 'communicatively non-dynamic', given the non-controversiality of the assumption that 'happenings' take place in politics.

- (23) WĒLL # – I don't THĪNK you know # that Tony Wedgwood Benn can seriously say that personalities – don't MÄTTER # – because I think they DÖ matter {TREMĒNDÖUSLY #} # in . [po?] (([ə])) PÖLITICS TÖDÄY # and especially in the politics of the LĒFT # – – what has HÄPPENED is # that

CĒRTAIN # – {DIVĪSIONS} of OPĪNION # have ARĪSEN # as they often DÖ # in the party of the LĒFT # . the Labour PÄRTY # (LL S.5.5.641–7)

(vi) *Tenor*

Tenor antecedents are 'located' in the speaker's thoughts or feelings. The speaker-centredness of the category is reflected in the frequency with which the highlighted element contains, in addition to a mental process verb or equivalent construction, a pronominal reference to the speaker. The speaker attributes to his/her previously unmentioned thoughts the status of 'shared background knowledge', thus presenting them as a relevant concern of the cooperative addressee. In (24) not only is the proposition 'I like something about it' syntactically backgrounded and presupposed, but within the proposition the affective mental process verb *like* is itself phonologically non-salient.

- (24) A oh NÖ # it's just the shop on the bridge just does everything CHEÄPER # I mean YÖU know # it's got EÖVERYTHING # . if ever you wanted some PÄRTS for {THIS #} # I would {ÄLWAYS} TRY it # ((2 sylls))
 (. . .)
 > A there's no doubt about CÄR batteries # anything like THÄT # ((it's[fæn])) the stuff they've got ĪN there #
 ? [m] #
 > A and . what Ī like about it ĪS # it's not all new branded STÜFF # . but it's not all RÜBBISH # (LL S.1.7.871–2)

Almost as common as the 'overtly speaker-centred' cases in the tenor category were those where a mental reaction or cognition, although certainly attributable to the speaker, is presented impersonally, as in (25).

- (25) The cause of that disaster may be revealed at the adjourned inquest. It took place on a part of the A20 that has a dual carriageway – which the people of Sidcup are still hoping will be extended into this urban district – so the need for a road improvement of that nature cannot be argued in this case.
What is alarming is not only that this sort of accident can still happen with dual carriageways, but that there could so easily have been other fatal accidents within the urban district over the week-end. (LOM B25.125–8)

(vi) *Mode*

(Mode antecedents are construable through a metalinguistic operation involving an interpretation or comment upon, reference to, clarification of, or remaking of a previous section of the discourse. It is easy for metalinguistic comments to be accepted as 'offstage' or 'background' material since they represent assumptions that are shared by cooperative discourse participants.) In (26) the givenness of the focally non-prominent items *you're saying (is)* may be argued to derive from such assumptions (though another possible source is the inferrable tie with *your point* in the previous tone unit).

- (26) B one can't start DEVELOPING this sort of THING # in a building where there's . not even the RIGHT sort of room # for ROUTINE Scottish teaching # AND [ɔ:m] # - f [ʔ] I want to draw attention to (THIS aspect) of what Jake has SAID # . I rudely INTERRUPTED you # .
 D it's ALL right TOM # -
 A come IN # - ((syll 2to3 sylls nice to SEE you #))
 D I've been teaching . I've been teaching over at [di] . ((ENG: NEERING department # 2 sylls))
 A (laughs --) come on IN # - LOOK # I I take [i] I take your POINT # WHAT you're saying is # we NEED # a {MULTIPLE} type of . I MEAN # IDEALLY we NEED # a multiple type BUILDING # (LL S.3.4,154-8)

Sometimes the metalinguistic notion that the speaker/writer presents as assumable by the addressee is more specific than merely 'saying' or 'meaning'. The theme of the pseudo-cleft in (27) is clearly more communicatively dynamic than that of the pseudo-cleft in (26). Nevertheless the information has an essentially 'offstage' character (one can readily assume that if a work is 'lively' then something will 'stand out' in it).

- (27) The views of M. Marand about the divine characters have been neglected and seem to me illuminating. In the same year as De Comus A Satan he published a pamphlet in English. The Effects of his Political Life on John Milton, concerned to show that a certain worldly-mindedness entered Milton's later poetry as a result of his rather sordid experience of government, politics and propaganda. *What chiefly stands out in this lively work, I think, is an accusation that Milton himself had smuggled into a later edition of Eikon Basilike the prayer, derived from*

Sidney's Arcadia, for which he then so resoundingly denounced King Charles in Eikonoclastes. (LOB J61,50-4)

* 6.1.3 *Basic pseudo-clefts and theme-rheme structure*

(Basic pseudo-clefts regularly exhibit a conflation of several distinct functions. The relative clause (fused, or together with preceding pro-form head), which functions as the grammatical subject, is mapped on to the semantic function of 'identified', the textual function of theme, and the logico-semantic function of presupposition.) (The bipartite nature of the construction is further reinforced intonationally by the (typical) realization of this set of functions as a single tone unit. What emerges from this complex of mappings is a characteristically 'interpersonal' flavour: a combination of the speaker-orientation of theme, and listener-orientation of givenness (supported by the backgrounding effect of syntactic dependency).)

In this section I shall concentrate upon the thematic structure of basic pseudo-clefts. The thematic flexibility of the construction is a feature which proves to be particularly relevant to attempts to explain the choice, in particular cases, of the basic pseudo-cleft in preference to other systemically agnate constructions. Sometimes this 'choice' is wide, sometimes limited. With pseudo-clefts belonging to the contextual categories of the taxonomy, the range of systemically agnate alternatives is limited, and particularly so with the field category. On the other hand there are generally more alternatives available with the co-textual categories. It is of interest to consider this finding in the light of the distribution of the different categories in speech and writing. Pseudo-clefts in writing have co-textual antecedents almost twice as often (64.6%) as contextual antecedents (35.4%); in speech the frequencies of co-textual antecedents (50.7%) and contextual antecedents (49.3%) are comparable. These figures are interpretable in terms of the suggestion, advanced here, that writing is richer in thematic resources, and that theme-rheme structure may represent its primary structural mode of textual organization. In speech, on the other hand, it may be that the primary mode of structural organization is given-new structure.

I shall begin by considering the thematic structure of pseudo-clefts in the contextual categories, beginning with field antecedence. The function of the *do*-type pseudo-clefts in this category, from a thematic point of view, is to enable sentence elements to be grouped in such a

way that verbal elements are assigned to both the theme and the rheme. The theme will contain the finite element(s): either pro-form *do* as the only finite element, or *pro-form do* preceded by one or more auxiliaries. The rheme will contain the head of the verb group (usually non-finite: infinitive, present participle or past participle). Some of the possibilities are represented in (28) below:

- (28) a. What Tom did was offer Sue a sherry.
 b. What Tom is doing is offering Sue a sherry.
 c. What Tom should have done was offer Sue a sherry.

Such pseudo-clefts thematize a nominalization consisting minimally of relative element + subject + verb group (with pro-form *do* as head). At the same time they rhematize the lexical verb denoting the process for which pro-verb *do* stands, optionally accompanied by further sentence elements. The performing of an action, pursuing of an activity, and the like is thus presented as both background and foreground. That a process was carried out, is being carried out, should have been carried out, and so forth, is presented as assumable knowledge by the speaker/writer, and then the specific nature of the process is presented as 'news'. (It may be noted that processes cannot normally be thematic in English, unless nominalized. The pseudo-cleft construction enables sentence elements to be grouped in such a way that a process, in the form of a cataphoric pro-verb, is made part of the theme.)

Consider example (21) again. The topic, or 'hyper-theme', of the discourse of which it forms part centres upon action – in particular the actions one needs to perform in order to accomplish the feat of eating cherries without using one's hands. In such a context it is not surprising to find a number of pseudo-clefts of the *do*-type of field-antecedence. In the case of (21) the theme of the corresponding non-cleft would be *you* ('you can do it only by sticking your tongue out ...'). In terms of the process-centred method of development of this discourse the non-clefts, with their actor-themes, would be less desirable than the pseudo-clefts, with their process-themes. The non-clefts fail, furthermore, to express the implicature that the action being described is the one that the speaker believes qualifies as the most appropriate method of performing the feat. The corresponding reversed pseudo-clefts ('to draw a line through there was what I was trying to do') would be superior to the non-clefts in their expression of exclusive identity, but less coherent than their basic counterparts in terms of theme-rheme structure. The highlighted element, which

represents the most communicatively dynamic side of the 'equation', would be too strongly foregrounded when presented as theme.

Finally, it may be noted that there is no cleft counterpart available to the pseudo-cleft with highlighted element functioning as complement to the pro-verb *do* ('it was to draw a line through there that I was trying to do'). The option available with pseudo-clefts of including the process in the highlighted section/identifier is generally not available for clefts.

It is the occurrence of pseudo-clefts where the theme is simply a reference to a 'happening', and every component of the happening is put into the rheme, that constitutes perhaps the strongest evidence for Halliday's (1985:43) claim that pseudo-clefts have evolved in English as a 'thematic resource'. The addressee is directed by the pro-verb towards the rheme for a specification of the process and the participant(s) involved in it. In some cases it is simply an element of mild assertiveness that is thematic, as in (29), where the meaning signalled by the choice of falling-rising nucleus could be glossed as 'actually'.

- (29) D [θi: ə] – the CHARACTERISTIC of our house # . is
 COFFEE cups #
 c [m] .
 D and what HÄPPENS is # that you may make coffee for SIX
 # – and put the empties in the SINK # (LL S.1.12.444–6)

Such an element might be represented alternatively (or, as well) by an adjunct, as in (22) above, or by a modal auxiliary, as in (30) below.

- (30) Man is the latest dominant type to arise in the evolution of this earth. There is no possibility of his dominant position in evolution being challenged by any existing type of creature, whether rat or ape or insect. *All that could happen to man (if he does not blow himself up with nuclear bombs or convert himself into a cancer of his planet by over-multiplication) is that he could transform himself as a whole species into something new.* (LOB G69.117–20)

Such *happen*-type pseudo-clefts are thus thematically similar to clefts such as *It wasn't that Tom offered Sue a sherry, it must have been that Tom offered Sue a sherry*: see further Section 6.3.4 below.

There is further evidence to support the suggestion that the explanation for this type of pseudo-cleft is essentially thematic. Consider (31):

- (31) They have developed a new system whereby completely untrained workers can be taught their trade by means of tape recordings and television.

What happens is that the unskilled worker is processed, by high-speed listening to recorded instructions on how to do the job coupled with explanatory OTV pictures, into becoming a highly skilled, obedient craftsman in no time at all. (LOB BOS,67-70)

If we were to insist upon a strictly formal notion of 'non-cleft counterpart', then the agnate non-cleft would be 'That the unskilled worker is processed by high-speed listening to recorded instructions . . . in no time at all happens'. A more acceptable non-cleft version would be one which did not include *happen*, namely: 'The unskilled worker is processed, by high-speed listening to recorded instructions . . . in no time at all.' This would seem to confirm that the role of the *pro-verb*, in the pseudo-cleft, is more to facilitate the thematic bracketing than to introduce ideational content (a nominalized clause could not be formed, to serve as theme, if there were no verb to serve as predicator for it).

Several typical examples from the remaining two contextual categories of the information classification – tenor and mode – will provide some idea of thematic limitations of these categories *vis-à-vis* the co-textual categories. Consider (32), from the tenor category:

- (32) (. . .) # . and very *OFTEN* # . [ði:] specialist *STUDY* # . [ði:] pursuit of the *DISCIPLINE* # is talked about especially by *NON-SPECIALISTS* # . as if it *WERE* # – a *NARROWING* process # *WHEREAS* # it is in fact . a *BROADENING* process # . but what worries *ME* is # . that . we often talk very *GLIBLY* # . as if . [ði] multiplying of *SUBJECTS* # is *SOMETHING* # that *CAN* # . [ə] work *AGAINST* # *NARROWNESS* # and *SPECIALIZATION* # (LL S 5.7. 844-51)

Here the proposition 'something worries me' is 'in the air' only to the extent that the speaker can reasonably expect a cooperative addressee to be concerned with his or her mental state. This proposition would be rhematic in the corresponding reversed pseudo-cleft ('that we often talk very glibly . . . is what worries me') and cleft ('it is that we often talk very glibly . . . that worries me'). Only in the extrapolated version of the non-cleft ('it worries me that we often talk glibly . . .') would the proposition remain thematic. Nevertheless there is a

clear communicative difference with the original: in the non-cleft the proposition is asserted (with *worries* carrying information focus in the unmarked instance), but in the original it is presupposed.

Consider finally (33), from the mode category:

- (33) These facts and figures are the measure of the problem facing the selectors at this moment, with Conroy injured and two other likely candidates for the forward line, internationals N.M. Forster and P.B. Austen, not available. They are also the answer to those critics of Conroy who complained that he slowed up the attack.

What they meant, of course, was that Conroy did not belong to the school that bash on regardless and hope for the best. (LOB AOS,165-6)

Here the proposition 'they meant something' is 'in the air' to the extent that the writer can expect the reader to infer that, in making their complaint, the critics referred to in the previous sentence were expressing a meaning. The reversed pseudo-cleft ('That Conroy did not belong . . . was what they meant') would not only be awkwardly front-heavy, but would assign informationally new status to material in the proposition 'they meant something'. The corresponding cleft is of dubious acceptability ('It was that Conroy did not belong . . . that they meant'). The agnate non-cleft ('They meant that Conroy did not . . .') is similar to the original in word order, but it asserts, rather than presupposes, the proposition that 'they meant something'.

With basic pseudo-clefts displaying co-textual antecedence, I have argued, discourse-coherent agnates are more readily available. Two examples will suffice to confirm this. The relative clause of the pseudo-cleft in (34) has an explicit antecedent in the previous sentence, and is thus appropriately presented as theme.

- (34) I think this attitude is short-sighted. First, nobody can transfer power, except in a purely legal sense. *What is transferred is legal authority.* (LOB G73,12-13)

In both the corresponding reversed pseudo-cleft ('Legal authority is what is transferred') and cleft ('It is legal authority that is transferred'), *legal authority*, the theme, would retain its status as new information (since the theme in reversed pseudo-clefts and clefts represents new information in the unmarked instance: see further Sections 6.2.1 and 6.3.1 below). The corresponding non-cleft ('Legal

authority is transferred') may cause temporary decoding difficulties for the reader, who may not realize immediately that the sentence has marked information focus.

The antecedent to the relative clause in (35) is the inference that poets, in pursuing their careers, are continually in search of something.

- (35) A gleam of gold shone in the front teeth of Burns Singer as he lit his fifth home-made cigarette. He said: 'Of course, Christopher believes that propaganda and politics are part of poetry.

'For me, it's different. It's almost like psychoanalysis. I'll do no work for weeks and then write solidly for 12 hours. I think *what I'm really seeking all the time is the source of Original Sin in myself.*' (LOB A19,155-6)

Agnate alternatives here would be less discourse-coherent than the original, but not remarkably so. One problem, for instance, with the reversed pseudo-cleft ('the source of Original Sin in myself is what I'm really seeking all the time') is the thematic foregrounding of the highlighted element, which would not be congruous with the method of development in the second paragraph (namely, references to the speaker Burns Singer, his working methods and objectives – as opposed to those of his fellow-poet). Comparable from a thematic point of view, but slightly preferable informationally, would be the corresponding cleft ('it is the source of Original Sin in myself that I'm really seeking all the time'). Here the local structure created by predication confers the status of unmarked newness upon the *source of Original Sin in myself*. In the reversed pseudo-cleft this constituent, in the absence of structural clues as to its informational status, might temporarily be misinterpreted as given. The corresponding non-cleft ('I'm really seeking all the time the source of Original Sin in myself') would have a similar word order to the original, but different thematic organization. Rather than a clausal theme, the non-cleft has simply the subject pronoun *I* as theme. It furthermore lacks the implication that the exclusive object of Burns Singer's search is the source of Original Sin in himself.

(To conclude, basic pseudo-clefts with contextual antecedents (which are more common in speech) have fewer acceptable agnate related constructions and may less readily be substituted by these in discourse, than basic pseudo-clefts, with co-textual antecedents (which are more common in writing). For further discussion of the

relationship between the thematic properties of basic pseudo-clefts, and their distribution in speech and writing, see Section 7.5 below.

6.2 REVERSED PSEUDO-CLEFTS

6.2.1 Reference of highlighted element

As Table 6.1 above indicates, the majority of reversed pseudo-clefts are realized as a single tone unit (289 of 351, or 82.3%). Of these 95.5% have nucleus placement in the rheme/relative clause, and of the total class of reversed pseudo-clefts, including those realized as more than one tone unit, 93.4% have nucleus placement in the theme/relative clause. (Despite these figures I shall regard reversed pseudo-clefts with a focal highlighted element as unmarked, on the grounds that they display, along with both clefts and basic pseudo-clefts, the unmarked correspondence between the identifier and new information.) The explanation for the figures is to be found in the nature of the element which regularly serves as highlighted element/ theme, demonstrative *this* and *that* (in 84.3% of cases in LL, and 77.1% of cases in LOB). In the vast majority of cases the demonstrative has anaphoric reference, and will thus naturally be non-focal, representing given information. Thus in the typical instance a reversed pseudo-cleft such as 'that's what I MEANT' will be unmarked in terms of an intonationally based description of information structure, even though marked in a broad structural sense.

Halliday (1967a:231) identifies three exceptions to the typical condition in which demonstrative *this* and *that* are non-focal because non-contrastively anaphoric. The demonstrative may be anaphoric but contrastive, exophoric, or cataphoric. In each of these functions, claims Halliday, the demonstrative will normally be informationally focal ('demonstratives are normally non-focal when anaphoric and focal otherwise', p.231). The word *normally* is important here, because LL contains several reversed pseudo-clefts with exophoric or cataphoric, but non-focal, demonstrative as identifier. Information focus is thus not fully determined by the reference function of the demonstrative. The typical pattern, however, is as represented in the following examples:

- (36) a. John's wife has left him and # *that's why he's UPSET* #
 b. Everyone knows that John's worried about his gambling debts, but they probably haven't heard the latest news: his wife has left him and # *THAT's why he's upset* #.

- c. I've had enough time now to choose the ring I'd like - #
THAT's the one I want #
 d. # *THIS is what he said* #: 'Any ring in this shop that takes
 your fancy will be yours'.

In (36a) *that* is anaphoric, and non-local. In (36b) the locus on *that* shows that the speaker wishes to contrast one reason for John's being upset against another. In (36c) *that* refers to a particular item in the context of situation that is being freshly identified. In (36d) *this* is local because cataphoric: it refers to material that, not having been introduced to the discourse as yet, is new.

In the typical case, as claimed above, the demonstrative is non-focal and anaphoric (as in (36a)). The type of anaphoricity exhibited by the demonstrative in such cases is usually of a special kind, termed 'extended text reference' by Halliday and Hasan (1976:66). Halliday and Hasan claim that extended text reference 'probably accounts for the majority of all instances of demonstratives in all except a few very specialised varieties of English' (p.66). This use of the demonstratives, together with the related use of *it* and *which*, represents in their view 'one of the major cohesive devices of the English language' (p.67).

'Extended text reference' is a term which actually covers two distinct processes. Halliday and Hasan (1976:52-3) call these respectively 'extended reference' and 'text reference'. Extended reference involves reference to a process, or complex phenomenon, or a sequence of these (for example, an eyewitness's remark: *It all happened so quickly*). It differs only in extent from usual instances of reference. Text reference differs in kind from usual instances of reference, and extended reference: the referent is transmuted into a fact or report. Halliday and Hasan give an example which illustrates the two types respectively (p.52):

The Queen said: 'Curtsey while you're thinking what to say. It (i.e. curtsey(ing) while you're thinking what to say) saves time.' Alice wondered a little at this, but she was too much in awe of the Queen to disbelieve it (i.e. the fact it saves time if one curtseys while thinking what to say).

Halliday and Hasan (1976:67) suggest that equative clauses ('pseudo-clefts', in our terminology) represent perhaps the most frequent form taken by extended text reference. The demonstrative 'provides the "given" element in the message and this then serves to identify some other element that is "new", by simply being equated with it'. As

Halliday (1967a:232) observes, the combination of (anaphoric) deixis with identification in these constructions is a particularly collective form of communication: a deictically recoverable stretch of discourse is represented as theme and offered as exclusively satisfying the 'definition' provided by the relative clause. Halliday and Hasan (1976) conclude their discussion of extended text reference and the pseudo-cleft construction with a claim whose validity is confirmed by the analysis of text examples below: 'Spoken English is typically held together by internal cross referencing of this kind, which combines powerful structure with great flexibility and freedom of movement' (p.67).

The figures for LL and LOB are displayed in Table 6.3. Extended reference and text reference, as this table shows, are restricted to *that*, *this*, *which* and *it*. The remaining, less frequently occurring items which serve as the highlighted element in reversed pseudo-clefts, typically have non-extended anaphoric reference. The various categories are exemplified below.

(i) *Exophoric highlighted element*

In (37) the demonstrative identifier *that* refers deictically to a building which is present in the environment in which the utterance is made.

- (37) Tommy pointed to a tall, important-looking building rising high above the other blocks of buildings. 'See right at the top, mister, where the curtains are - *that's where we live*.' (LOB P28,128)

(ii) *Cataphoric highlighted element*

Clear cases of reversed pseudo-clefts with cataphoric identifier are extremely rare. An example is (38):

- (38) Yet if, because of his broad generalizations and his imperviousness to tinsel compliments, we used to think him unworldly, we were at once overestimating and underestimating him. For he has shown - and it is *why the Rede Lecture has such an authoritative ring* - a fine grasp of the realities of power. (LOB G34.89-90)

(iii) *Anaphoric, but contrastive, highlighted element*

This category is different from the other purely phoric categories insofar as it is defined in terms of intonation as well as phoricity (it is thus not represented in Table 6.3: see Note 4). The category thus overlaps with the three anaphoric categories described below.

There are fifteen cases in the spoken data of reversed pseudo-clefts with a demonstrative highlighted element that is both focal and

Table 6.3 Reversed pseudo-cleft: reference of highlighted elements*

Reference	Corpus	that	this	which	it	these	those	Per ¹ pron. ²	NP ³	VP	PP	Adj	Adv	Total ⁴
Exophoric	LL	2	1					3						6 (0.1)
	LOB	10	9					1					1	21 (0.2)
Cataphoric	LL			1										1 (0.02)
	LOB				1									3 (0.03)
An extended	LL	14	1	1	4	1	1	10	8		1			41 (0.9)
	LOB	12	2	2	4			2	19	2		1		44 (0.4)
Extended	LL	91	33	15	1									140 (3.2)
	LOB	31	13											44 (0.4)
Text	LL	118	36	3	1									158 (3.6)
	LOB	76	13	5	2	2			1					99 (1.0)
Total	LL	225	72	19	6	1	1	13	8		1			346 (8.0)
	LOB	130	38	7	7	2		3	20	2	(0.02)			211 (2.0)
			(1.3)	(0.4)	(0.1)	(0.02)		(0.03)	(0.2)	(0.02)		(0.01)	(0.01)	

Notes

* Bracketed figures represent frequencies per 10,000 words.

¹ There were 23 tokens in LL with nucleus placement in the rheme (see Table 6.1 above) comprising 3 with exophoric highlighted elements (while another 3 exophoric highlighted elements were non-local), 1 with cataphoric highlighted element, 15 with contrastively new highlighted element (see Note 4 below) and 5 with 'freshly' new highlighted element (see Note 3 below).

² Tokens been emerged separately from the other personal pronouns (i, you, we, and so on) because it is the only member of the class that can have reference.

³ In addition to the 15 tokens in LL with an anaphoric noun phrase as highlighted element, there were 5 with a 'new' theme (making a grand total of 351 reversed pseudo-clefts in LL).

⁴ 15 tokens in the anaphoric categories of LL were contrastively new, and are not represented in the table.

clearly anaphoric, and that must therefore be interpreted contrastively. For example:

- (39) A (. . .) there's a man called Aldo CALIFARO # . you ever hear of HIM # .
 B WELL # I've read one or two things he's WRITTEN # (. . .) # - so he might come rather EXPENSIVE # he's earning nineteen thousand DOLLARS # -
 A he's[a] he [d] he ((descants)) on the question of money rather [o:] . in DETAIL #
 B YEAH # he well he says two THINGS # A # . he's quite prepared to take a DROP # and then B # . at the END # PS # . what are you going to OFFER me # (. . .)
 C I think he he [a] wants to be near the British MUSEUM #
 B YEAH #
 C to ((go in for)) Spenser VARIORUM # - do you think THAT'S what it is # but it's it's not an ignoble . WISH # (LL S.2.6.664)

In (39) the presuppositional set is implicitly derivable. Speaker C is a member of a selection committee which is screening applicants for an academic position. The applicant in question, Aldo Califaro, has indicated a willingness to accept a reduction in salary in the position applied for, so it can be assumed that he has other-than-mercenary motives in applying. The *it* of the relative clause presumably refers to the applicant's motives, these constituting the presuppositional set, while *that* identifies Mr Califaro's putative desire to be near the British Museum as the set-member to which C wishes to attach prominence.

(iv) Anaphoric (non-extended) highlighted element

This category is represented by a wider variety of grammatical items than the other categories of phoricity. Examples from LL and LOB follow:

- (40) (a) (. . .) black HANDS # - [ði ə] the ends of the toes are ALSO black # and so is the TAIL tip # . and this is ONE of the FEATURES # {OF} [ði:] red neck((ed)) WALLABY # which is . what this one IS # (LL S.10.8, 690)
 (b) I came fresh from two years of research, spent among the high stone pillars and solemnly dedicated atmosphere of the

British Museum. I am essentially a provincial lad, lost in the vast, unwieldy city of London, and *the British Museum was the only place I knew*. (LOB K26,22)

(v) *Anaphoric (extended) highlighted element*

As Table 6.3 indicates, reversed pseudo-clefts with extended anaphoric highlighted element are outnumbered by those with text-anaphoric highlighted element in both the spoken and written data. The ratios differ, however, with text reference being comparatively more popular than extended reference in writing than it is in speech. This finding is compatible with the notion discussed in Section 1.4 above that speech is characterized generally by an orientation towards action, writing by an orientation towards reflection.

In (41) the complex phenomenon referred to by *this* in the first pseudo-cleft is the practice in legal work of consulting reference works in regular succession, while that in the second pseudo-cleft refers to methods in the study of English where, by contrast with the legal practice, a number of books are said to be referred to simultaneously.

- (41) >A can I put it **THIS** way # . it seems to MĒ # that in LĀW #
 . there's a PRŌBLEM # .[ə:] I.ĀWYERS when they
 work # have to be surrounded by LĀW reports # and
 they put down a VŌLUME **THERE** # they LŌOK it ŪP
 # . with in -- QUARTER of an hour # they've got the
 STUFF ŌUT of that # they PŪT the thiug BĀCK # .
 they want ANŌTHER # . because there's another one
THERE # . *this is what they're doing the whole TIME* # .
 now in {ENGLISH} this doesn't seem to me to be the
 PRŌBLEM # . in ENGLISH # you may have to WŌRK
 # from four or five or six books SIMULTĀNEOUSLY #
 . spread round a TĀBLE on you # .[ərau] RŌUND you
 # . but . there you ĀRE # for the next seven HŌURS # .
 [ə:m] ((2to3 sylls))
 [ə:] (- laughs)
 VAR (--- laugh)
 A is this FĀIR # . *is this what HĀPPENS in English* #
 (LL S.3.3,151,163)

(vi) *Anaphoric (text) highlighted element*

Reversed pseudo-clefts with a text-referential highlighted element were noted to function as important cohesive devices in all varieties

of spoken and written discourse. In (42) the complex phenomenon that is transmuted into a fact is the unsatisfactory development of the coast referred to by Speaker c.

- (42) c yeah but I mean that that part of the coast isn't very well
 developed is it
 A [m] # - . NŌ # ((this)) *this is what's so* [əm]
 c ((several sylls interesting)) [m] .
 >A so ATTRACTIVE about it # (LL S.2.13,846-7)

6.2.2 *Discourse functions of reversed pseudo-clefts*

By contrast with the highlighted element of basic pseudo-clefts, which regularly represents new information, that of reversed pseudo-clefts typically represents given information (the small set of exceptions includes freshly new noun phrases, anaphoric but contrastively new demonstratives, exophoric and cataphoric demonstratives and pronouns).

The relative clause of the reversed pseudo-cleft typically contains an intonation nucleus. This follows from the fact that the relative clause functions as rheme within a sentence that is regularly realized as a single tone unit (so the nucleus falls in its unmarked position at, or near, the end of the tone unit). The relative clause of the reversed pseudo-cleft is thus, in terms of intonation, not unlike that of the basic pseudo-cleft. The latter also typically contains a nucleus, in this case because the relative clause is usually realized as a separate tone unit (see Table 6.1 above). Because the relative clause is rhematic in the reversed pseudo-cleft, it is more likely to contain new information than that of the basic pseudo-cleft, which is thematic. There are nevertheless two factors (see further Section 6.1.1 above) which operate to attenuate the newness of the item or items so marked within the relative clause. Firstly, the dependent status of the relative clause has a backgrounding effect, conferring a non-controversial character upon the information contained in it. Secondly, the relative clause represents a presupposition, a proposition whose truth is assumed by the speaker.

If, as the preceding discussion suggests, the reversed pseudo-cleft may contain little dynamically new information overall, then what is the communicative function of the construction? One clue is provided by the distributional patterning of reversed pseudo-clefts in texts. Typically they occur at the ends of stages, or sections, in a discourse, serving a role that might be characterized as 'summative'. There are

two reasons why reversed pseudo-clefts should be suited to this role. The first reason is that 'endings' are appropriately signalled by sentences of low informational content (which, depending on the register, may take the form of generalizations, clichés, explicit repetitions, and so forth). Stubbs (1983:24) has observed that the ending of stories in casual conversation is often signalled by the use of a 'cliché-cum-proverb with little informational content'. Interestingly, of the four examples he gives (*Still, that's life*; *Well, that's the way it goes*; *But something may turn up – you never know*; *Still, we may as well hope for the best*), two contain the same item – demonstrative *that* – which serves as highlighted element in 62 per cent of reversed pseudo-cleft tokens in the present corpus, the second example being in fact a reversed pseudo-cleft. Stubbs offers the following explanation for this practice (p.24):

Such utterances with little significant propositional content, provide no new information which can serve as a resource for further talk, and can therefore serve as endings.

Stubbs further notes that repetition of whole phrases is also a marker of endings. Such repetition, as the discussion of corpus examples below will confirm, is a feature of many reversed pseudo-clefts.

The second reason why reversed pseudo-clefts should be suited to a 'stage-ending' role is their operation as 'internal referencing' devices. Typically, we have seen, the construction serves to equate a deictically-referred-to stretch of previous discourse with an assumable proposition which either specifies or exclusively defines it. It is the equative relationship itself which may provide the primary informational contribution of the construction to the discourse. In other words the message encoded by the speaker through the choice of the reversed pseudo-cleft construction can be glossed as something like: 'I conclude by drawing your attention to the identity of *x*, which I have been talking about, and *y*, whose evidentness I am assuming in this context.' The reversed pseudo-cleft thus offers a means of 'pulling the threads of the discourse together'. And it may do this in such a way that little or no informational content is introduced, and which may consequently border upon the platitudinous or clichéic.

In the discussion that follows several corpus examples are analysed, with particular attention being paid to their communicative and text-structuring roles. I shall begin with spoken examples from LL, and then consider written examples from LOB.

Each of the three reversed pseudo-clefts which follow one another in rapid succession in (43) has text-referential *that* as highlighted

element. Each demonstrative in turn refers to a slightly larger section of the preceding discourse and, at the same time, the relative clauses become successively less specific. The net effect is one of decreasing informational content, gradually bringing this section of the discourse to a close. At the same time the three reversed pseudo-clefts draw together what might otherwise have been disparate threads of the discourse into a tightly knit passage of cause and explanation.

- (43) M well now that you can stand back and look at Ireland Kevin what do you think of the mess over there –
 A well there doesn't seem to be much way OUT of it REALLY # everyone – – . seems to be very DEPRESSED # I mean the I think the GREAT tragedy # was the fact that
 f [m] yes [m]
 > A SUNNING SUNNINGDALE failed # , cos it so NEARLY # SUCCEEDED # . . ((I mean it)) when it came to the CRUNCH you see it was [o] # , (ENGLAND'S) trouble over the MINERS # took PRECEDENCE over # . Ireland's . problems with SUNNINGDALE # and that's why the election was called at that TIME # – and that's why Sunningdale . FAILED # . I mean I may have OVERSIMPLIFIED it # but I think LARGELY # ((that)) that's what HAPPENED # (LL S.2.8,13,14,17)

In the first reversed pseudo-cleft in (43) *that* is referential to '(the fact that) England's troubles over the miners took precedence over Ireland's problems with Sunningdale'. The *why*-clause represents assumable, background knowledge for the interlocutors (hence the definite article *the* before *election* and deictic demonstrative *that* before *time*). The reversed pseudo-cleft relates this (background) information to the statement about England's difficulties with her miners. The latter is exclusively identified as the reason for the calling of the election. In the second reversed pseudo-cleft the co-textual 'snowball' to which the thematic demonstrative refers has increased in size. It consists of both the 'effect' and 'cause' represented in the first pseudo-cleft ('that the election was called at that time because England's troubles over the miners took precedence over Ireland's problems with Sunningdale'). These are exclusively equated (as a reason) with information whose status is explicitly given (*Sunningdale failed*). In the first two reversed pseudo-clefts, then, it is the equative relationship itself, with its added circumstantial component (that 'known fact A' constitutes a reason for 'known fact B'), that makes

the most dynamic communicative contribution to the discourse. The final reversed pseudo-cleft (*that's what happened*) clearly has a text-structuring function. It has a confirmatory tone, and signals through its low informational content that Speaker A has completed his argument and his turn. The pseudo-cleft reaffirms the validity of the causal chain identified by the speaker ('Sunningdale failed, because the election was held at that time, because England's problems . . .') simply by restating its occurrence ('this chain of events happened').

In (44), as in (43), a series of reversed pseudo-clefts is used to cross-reference a section of a conversation in which the speakers seek to verify the status of a piece of gossip about a mutual acquaintance's husband.

- (44) A (. . .) # --- OH # JENNIFER came to lunch last WEEK #
 C ((bless my SOUL # - when did we LAST HEAR of her #))
 b ((gracious))
 A with her HUSBAND # did you KNOW she ((got)) married #
 C NO #
 A YES #
 C Pen I NEVER # - IM{AGINED} in my WILDEST DREAMS # ((4 sylls. and)) to WHO # in GOD'S name #.
 A (laughs) he's a very wild and woolly IRISHMAN # ((a)) WIDOWER # - at present UNEMPLOYED # BE{CAUSE he } SAYS # he ted a STRIKE #.
 C oh GOD # GOD # GOD # {oh} #
 A he is I THINK a ---- CRANK #
 C why did that get him UNEMPLOYED # -
 A well HE says # that he was then BLACKLISTED # and no employer will HAVE him # - but [ə] - I mean he MUST get C in IRELAND # [wə]
 A NO # . in ÉALING # -- or ACTON or SOMEWHERE #
 C you can't SÁCK a person for leading a STRIKE #
 A for SÁMPS # well he was apparently -- WELL # *this is what he SAYS ANYWAY #*
 (C again queries whether leading a strike would be sufficient grounds for a worker's dismissal, and acknowledges that an employer may offer this as a spurious justification, while concealing another motive . . .)
 A you'd { PROBABLY } have some OTHER REASON # . I don't KNOW # anyway he feels DISCRIMINATED against # *this is all I can TELL you #*

(C reiterates the difficulties of sacking people in England. A rejoins by observing that he is merely reporting what Jennifer's husband has claimed . . .)

A and he was perhaps sacked for something ELSE # - but I do not KNOW # or made REDUNDANT or something # - ANYWAY # *this is what he SAYS #*. (LL S.2.14.782,791, 806)

All three reversed pseudo-clefts here affirm Speaker A's incomplete knowledge of the circumstances of the 'wild and woolly' Irishman's dismissal. A's listeners question the plausibility of her explanation of the dismissal, and she rejoins by indicating that she is merely repeating the Irishman's claim. The function of the adjunct *anyway* in the proximity of the first and last pseudo-clefts is to convey an element of defensiveness, as A attempts to vindicate her role as bearer of 'second-hand news'. In each case the highlighted demonstrative *this* is text referential, representing the claims attributed to the Irishman that A is reporting to her interlocutors.

The first example establishes, through the rising-falling nucleus, a contrast between the Irishman's reported claim, and whatever the truth of the matter might have been. A similar contrast is suggested in the second and third examples, but in the absence of the falling-rising nucleus the opposition is perhaps less salient. Once again the position of the reversed pseudo-cleft (turn-final, in each case) is one that might be predicted from its confirmatory, summarizing function.

The extract in (45) is from an article entitled 'The Church and politics', which describes the conflicting interests of the Church and the State. It includes a true story about a Civil Servant by the name of John Williamson, whose life epitomizes this conflict. At first John attempts to keep away from politics, fearing that involvement would hinder him in his occupation. However, a series of minor incidents occurs which break down his resolve not to meddle in politics, culminating in a confrontation with the Local Education Authority. The extract in (45) describes the final incident, which brings matters to a head.

- (45) What brought things to a head? It happened this way. Although John wasn't a Catholic, his wife and children were. Marie, his eldest girl, won a scholarship. So, of course, he put her name down for the Convent of the Sacred Heart. It's true that the convent was in the next town, three miles away. But Marie was a big, strong girl. Half-an-hour's journey wouldn't do her any harm.

But what happened? He had a letter from the Local Education Authority telling him that Marie could not go to the Convent School. She would have to go to the Municipal High School. He wrote back, thinking there had been some mistake. He pointed out that although he wasn't a Catholic he'd promised to bring the children up Catholics. So, of course, his girl must go to the convent. Back came the reply by return post. High School or nothing, if he didn't send her on the first day of term they would prosecute.

That's why John started meddling in politics.
(LOB D16.121)

Here the highlighted item *that*, serving as unmarked theme, refers immediately to the incident with the Local Education Authority, but beyond this its referent is the accumulated co-text detailing John's increasing involvement in politics. Not only does *that* represent given information, through its (extended) anaphoric reference, but also the proposition 'John started meddling in politics' is very low in newsworthiness (the phrase 'meddle in politics' has been used previously in the article and, in any case, the proposition is readily deducible from the anecdote being narrated). What is being presented as news (though it is a confirmation of something that the reader has already been told, rather than 'fresh news') is that the incident(s) referred to, and John's political involvement, are related to one another causally.

Example (46) is from a love story in a women's magazine.

- (46) It was one of the strangest courtships I'd ever known – it held society gossips by the ears all that Season.

A pretty case for Freud.
BY ALEC WAUGH
ILLUSTRATED BY BOB PEAKE

I noticed him in the first place because he was the only other person in the pavilion wearing a silk hat.

(...)

I was conscious not of a handsome face, but of a new person: of someone who was masterful but unworlily; practical but inexperienced; masculine but with that look of anticipation, of waiting to be fulfilled that you expect to find in a young girl; a combination of characteristics so self-contradictory that the

obvious corollary to their catalogue would be: 'What a mass of complexes. A pretty case for Freud.' *That was what you would have expected.*

He wasn't, though. He was of a piece, without self-consciousness: the kind of man who does not know what the word shyness means. (LOB P19.57-8)

The demonstrative identifier *that* is here text-anaphoric to the immediately preceding quotation, which represents a deduction from the 'external evidence' (the man's appearance contains a number of apparently self-contradictory features). The proposition with which this 'report' is identified, 'you would have expected x', is 'implicitly given' (it is inferable from *obvious corollary*, and also picks up *that you would expect to find* earlier in the text). The reversed pseudo-cleft thus neatly rounds off the paragraph. And yet it contains, at the same time, a forward-looking element. As is so often the case with verb groups expressing counter-factuality, there is a residual element of counter-expectedness. The pseudo-cleft thus both internally references and concludes one paragraph, but links it cohesively with the next.

6.2.3 Reversed pseudo-clefts and theme-rheme structure

Reversed pseudo-clefts are, like basic pseudo-clefts, thematically unmarked insofar as they display the mapping of theme and subject that is normal for declarative sentences. In the reversed construction, however, it is the identifier rather than identified element that is thematic. The identified element (logico-semantically presupposed and syntactically backgrounded, as in the basic construction) is rhematic. The characteristically 'interpersonal' flavour of the basic pseudo-cleft, deriving largely from the association of theme and presupposition, is absent from the reversed pseudo-cleft, where these functions are realized separately.

Reversed pseudo-clefts, as we have seen in the previous section, typically have as identifier an 'inherently given' item. Such an item is most appropriately represented as the theme, which may be glossed as 'what I have been talking about until now, namely x.y.z...'. It follows that the effect of 'de-reversing' a reversed pseudo-cleft will typically be a combination of 'inherent givenness' and (insofar as the highlighted element will now be rhematic) an expectation of 'newness'. Such a combination is not of course impossible: it gives rise to a meaning of contrastiveness. However, a contrastive meaning will not

be appropriate in most cases. The highlighted element is generally used to refer back to a stretch of previous text, not in order to newly identify it, but simply in order to indicate that it is recoverable.

As with basic pseudo-clefts, so with reversed pseudo-clefts, there is much variation in the availability of agnate alternatives. Consider (47):

- (47) b (. . . most people have . . .) got romantic conceptions of MARRIAGE # and they rush INTO it # and they don't THINK
 a [m] #
 b about it PARTICULARLY # - - and[o:] - I've certainly learnt a LÖT by my experience # I THINK
 a YES #
 b I have # ANYWAY # - .
 a ((well)) we see it a lot at SCHOOL # if you KNOW ((sort of)) # . BOYS # . in the SIXTH form # - round about seventeen or EIGHTEEN # . and they become ENGAGED # . *that's what happened about two YEARS ago* # there was a whole sort of RÄSH # . nearly everybody in the SIXTH form # seemed to be getting ENGAGED and #
 (LL S.5.10,188)

Here the basic pseudo-cleft, cleft and non-cleft agnates are all structurally well-formed (*what happened about two years ago is that; it's that that happened about two years ago; that happened about two years ago*). However, none of these, if substituted for the original, would be as textually coherent. The basic pseudo-cleft has a thematic structure that is inappropriate in the context of the two 'terms' in the equation represented by the pseudo-cleft structure, it is clearly *that* rather than the relative clause which is the least communicatively dynamic, and which is by definition 'in the air'. The rhematizing of *that* creates an expectation of focal marking. This being so, the inherent givenness of the demonstrative demands that it be interpreted as either contrastively anaphoric, cataphoric or exophoric. But each such interpretation is out of place in the present context: the function of the demonstrative here is clearly (non-contrastively) extended referential. Similarly, the corresponding cleft would exhibit an information structure that would be contrary to expectation (insofar as the predicated element of the cleft sentence is focal in the unmarked instance). The non-cleft version would not be malcohesive (its theme-rheme structure is virtually identical to that of the

reversed pseudo-cleft). Nevertheless the contribution that it would make to the discourse would differ subtly from that made by the original. Whereas the pseudo-cleft asserts that an anaphorically-referred-to section of text is to be identified with a presupposition ('something happened about two years ago'), the non-cleft sentence might be characterized as 'all assertion and no presupposition' (and thus would give the impression of making a more dynamic contribution to the discourse).

The next example comes from a written text from the 'Popular lore' category of LOB.

- (48) That this was the case is confirmed both by the testimony of the earliest references to Robin Hood in the chronicles, and by the consistently favourable attitude of the outlaws of the story towards the poorer classes. The outlaws were not always poor men, but the poor man did not demand that. He demanded kindness, good lordship to engage his fidelity, and *this is what the outlaw gave*. It is the theme of Robin Hood's famous advice:

But loke ye do no husbonde harm,
 That tilleth with his ploughe.

(LOB F27,151)

The inherent givenness of the demonstrative makes it, as always, appropriate as an informationally unmarked theme. There is, by contrast, an element of (semi-) newness in the relative clause (it is inferable, but perhaps not entirely predictable, from the outlaws' 'consistently favourable attitude . . . towards the poorer classes', that they would meet rather than reject the demands of 'the poor man'). Thus the corresponding basic pseudo-cleft (*what the outlaw gave is this*) would be less appropriate thematically. It would, furthermore, misleadingly suggest to the reader that *this* is focal, and is thus to be interpreted as other-than-non-contrastively anaphoric. The corresponding cleft sentence (*it is this that the outlaw gave*) retains the same order of elements (those with ideational function) as the original, differing thematically in the textual prominence given to the demonstrative through predication. Like the basic pseudo-cleft it also invites a reading, unwarranted in this context, in which *this* is nuclear. There is a similar problem with the agnate non-cleft with fronted theme (*this the outlaw gave*). The thematically unmarked non-cleft version (*the outlaw gave this*) is slightly less textually suitable because its theme, *the outlaw*, is inconsistent with the method of development

in the final lines of the extract – namely the virtues on which the harmonious relationship between outlaws and poor men rested.

To conclude, I have argued that the reversed pseudo-cleft typically functions as a text-structuring device. Its highlighted element is normally a demonstrative or pronoun with extended text reference which, as theme, represents the speaker's point of departure. Consistent with the text-structuring role of the construction, the theme typically does not draw any attention to itself: it is structurally unmarked (as subject), and informationally unmarked (except on the rare occasions when it is contrastively, situationally or cataphorically new). The equation of the element(s) functioning as the theme with the element(s) functioning as the rheme typically has an internal referencing role. The discussion in the present section has indicated that reversed pseudo-clefts are by no means merely 'free variants' of basic pseudo-clefts (or, of other agnate constructions). When the reversed pseudo-cleft is studied in natural discourse, its specialized role, deriving from the combination of the equative structure with extended text anaphora, becomes apparent. Exploration of the differing communicative roles of the reversed pseudo-cleft and its agnates thus provides supporting evidence for Bolinger's (1972b:71) general claim that:

There are situations where the speaker is constrained by a grammatical rule, and there are situations where he chooses according to his meaning . . . ; but there are no situations in the system where 'it makes no difference' which way you go . . . This is just another way of saying that every contrast a language permits to survive is relevant, some time or another.

6.3 CLEFTS

6.3.1 Intonation and information

In cleft constructions, I have argued, the unmarked location of information focus is, as in pseudo-clefts, on the highlighted element. There are three main reasons why clefts with focal highlighted element should be regarded as unmarked. Firstly, within the local structure (*it + be + . . .*) created through predication in the cleft sentence, the intonation nucleus is in its unmarked position at the end. Secondly, if the highlighted element is focal there is an unmarked correspondence between the element functioning as the 'identifier' and new information. Thirdly, one of the functions of the

cleft is to give thematic prominence to the highlighted element. It follows that in clefts (as Halliday 1967a:237) observes) 'since thematic prominence is a form of "new" information, the predicated element carries the unmarked information focus.'

The unmarked pattern in clefts is, as in the case of reversed pseudo-clefts, commonly overridden. As with reversed pseudo-clefts (though much less frequently) a non-contrastive anaphoric item (demonstrative, pronoun, 'scene setting' adverbial, and so on) will be selected as highlighted element. Halliday (1967a:237) summarizes the informational possibilities of clefts thus:

There is the usual pattern of unmarked association: a prominent theme will normally be new, but it may in any instance be given, including all instances where it is anaphoric unless also contrastive.

There is a widespread misconception that the highlighted element of clefts always carries contrastive focus (for example Quirk *et al.* 1972:951, Harries-Delisle 1978). However, while cleft sentences are frequently associated with a formulation of contrast, the highlighted element may not be contrastively focal (that is, incorporating placement of the intonation nucleus – and thus definition of the culmination of the new information – upon some item other than the last functional element of clause structure in the tone/information unit). In (50), for instance.

- (49) He'd rushed to the surgery and was breathing heavily. But it wasn't *that* which disturbed me. It was the time he took to recover. (LOB F33,102,103)

a contrast is presented between the patient's 'heavy breathing' and 'recovery time'. And this contrast will be established whether or not the highlighted element in the first cleft sentence, *that*, carries an intonation nucleus. It is difficult in fact to predict the way a reader is likely to interpret the information structure of this sentence, given the competing pressures of unmarked focal-marking of the identifier on the one hand, and the typical non-focality of anaphoric demonstratives on the other. If the demonstrative is read as focal, then, being anaphoric, it can only be interpreted contrastively. An example from the spoken corpus in which there is contrast, but not contrastive focus, is the first cleft in (50).

- (50) (. . .) and I want *HERE* CHAIRMAN # – to *ASK* # if you could *TELL* us # – – what [*ði*] notion *IS* about college at the

MÓMENT # about what we MIGHT CÁLL #
INTERDISCIPLINARY. ([æ]) { STUDIES # } # . Scottish
could you see be SEEN # as a service INDUSTRY ((in))
{ some . RESPECTS # } # . [ə:m] it's not that *aspect* 'in
ÁFTER # it's the fact that there's a fair NÚMBER of us # in
this SCÓTTISH department # . WHÓ # . might just as WÉLL
be doing the things we're very GÓOD at # in other sorts of
DEPARTMENTS # (LL S.3.4.863)

Here *that aspect* contrasts with 'the fact that there's a fair number of us . . .', but it is not contrastively focal.

There are further potential sources for the misconception that the highlighted element of clefts always carries contrastive focus. One is the exclusiveness implicature which I have argued above is a feature of both cleft and pseudo-cleft constructions and derives from their identifying structure. Another is the semantic feature which has been noted by Halliday (1967a:236) to derive from the mapping relationship between identifier and theme. Halliday observes, in the course of a discussion of the sentence, *it was John who broke the window*, that:

Structurally predication maps the function of identifier on to that of theme, giving explicit prominence to the theme by exclusion: 'John and nobody else' is under consideration.'

Some of the claims that have been made in this section can be quantitatively evaluated by examining the intonational characteristics of cleft sentences in LL. The relevant statistics are presented in Table 6.4. Of the 187 cleft sentences in LL, 74 (or 39.5%) are realized as a single tone unit, and 113 (or 60.4%) as two or more tone units. Of the 74 realized as a single tone unit, 34 or (45.9%) are structurally unmarked (that is, the highlighted element was focal), and 40 (or 50.4%) are structurally marked. In both cases the falling nucleus is by far the most popular (accounting for 61.7% of the unmarked instances, and 65.0% of the marked). Of the 113 realized as two or more tone units 91 (or 80.5%) are structurally unmarked and 22 (or 19.4%) are structurally marked. Of the 91 structurally unmarked tokens, 63 (or 69.2%) exhibit 'unmarked tonality', with the superordinate clause realized as a single tone unit (with, in 87.1% of cases, unmarked nucleus placement on the last functional element of clause structure in the information unit, and with the falling nucleus again being the most popular). It follows from these figures that the prototypical cleft sentence is that realized as two (or more) tone units, with

Table 6.4 Intonational features of clefts in LL

<i>No. of tone units realizing cleft</i>	<i>Theme type</i>	<i>Tonicity/tonality features</i>	<i>Structural unmarkedness/ markedness</i>	<i>Subtotals</i>	<i>Total</i>
one	topical	focal high-lighted element	marked	34	74
		non-focal high-lighted element	marked	30	
	non-topical	non-focal high-lighted element	unmarked	10	
more than one	topical	superord. cl. one t.u.	unmarked	28	187
		superord. cl. = one t.u.	unmarked	53	
	non-topical	superord. cl. = one t.u.	unmarked	10	113
		superord. cl. one t.u.	marked	22	

the first tone unit mapped on to the superordinate clause and exhibiting unmarked tonicity. An example is (51):

(51) SÔ # although [2] I could imagine that we { CÒULD } . [ə:m].
on our joint SALARY # get perhaps quite a a high
MORTGAGE # its the paying it back at the BEGINNING #,
that's going to be DIFFICULT # (LL S.8.1p.1082-3)

Considerably rarer are examples of the type given in (52), with similar tonality, but marked tonicity in the first tone unit.

- (52) well it's . mainly a LÄRKISH course # – that [ə] we're MÄKING # (LL S.9.21,638-9)

Of the 113 clefts realized as two or more tone units, 53 are of the type illustrated in (51) (accounting for 46 tokens), and in (52) (accounting for 7 tokens). The other 60 are divided almost evenly between those with superordinate clause realized by two or more tone units (28 tokens) as in (53), and those without a topical theme and thus typically with superordinate clause not realized as a separate tone unit (32 tokens), as in (54).

- (53) A (. .) # . [əm] – – well { YÉS } there ÄRE techniques to LEÄRN # . but it's – very MÜCH the #
? [nɪ] # .
APPLICÄTION # or the CRITICISM # ÖF # the very techniques THEMSELVES # which is the . IMPÖRTANT thing # (LL S.3.3,1256-62)
- (54) (. .) # . it's jst that his WRITING # ((suggested this sort of RING # (LL S.1.6,1096-7)

Although with clefts realized as more than one tone unit there is a typical form, as exemplified in (51), this is not the case with clefts realized as a single tone unit. Slightly more than half in the latter category are structurally marked (30 tokens with topical theme, as in (55), and 10 tokens with non-topical theme, as in (56)). The remainder (34 tokens) are structurally unmarked, as in (57).

- (55) # it was Norma Harley that did most of the TÄLKING # (I.L. S.3.2,1035)
- (56) # [ə] it's not that Mervyn's [ə? ə?] TÖTÄLLY unreliable # (LL S.2.6,1119)
- (57) (. coughs) WÉLL # she must have KNÖWN about it # and and . ((it was)) SHÉ who burnt them attempted to . she attempted to burn the BÖÖKS # (LL S.1.13,702)

6.3.2 Clefts and information structure

In Section 5.3.5 above I discussed the information structure of cleft constructions, and proposed a classification comprising three types (each involving four combinations of categories of informativity: new, contrastive, given and inferable). The discussion of corpus examples which follows is organized in terms of the twelve cleft subtypes thus defined.

(i) Type 1 clefts

The vast majority of Type 1 clefts in LL are realized as a single tone unit, with focal highlighted element, and sometimes with ellipsis of the relative clause. A smaller number of Type 1 clefts are realized as more than one tone unit. In these cases any information presented intonationally as 'news' in the relative clause, is of very low communicative dynamism, with intonational signals assuming a role of secondary importance to textual indications of informativity. Type 1 clefts are relatively common in I.OB, a not unexpected finding in view of the latent preference in writing for intonational unmarkedness (see further Section 7.4).

(a) New-given

The two clefts in (58) participate in the formulation of a contrast. (The explicit parallel between *that aspect* and *the fact that . . . departments* warrants the treatment of the latter as a cleft with ellipsed relative clause rather than a simple equative with referential *it*.) It is the second of these clefts, as italicized, that belongs in the present subcategory. The ellipsed clause is explicitly retrievable (*I'm after*); the highlighted element, by dint of its very length, could not but represent 'fresh' news.

- (58) (. .) # and I I want HÈRE CHÄIRMAN # – to ÄSK # if you could TÉLL us # – – what[ä:] . notion IS about college at the MÖMENT # about what we MIGHT CALL # INTERDISCIPLINARY. (([ä])) { STÜDIES # } # . Scottish could you see be SÈEN # as a service INDÜSTRY(([ä])) some . RESPECTS # # . [ə:m] it's not that aspect I'm ÄFTER # it's the fact that there's a fair NÜMBER of us # in this SCÖTTISH department # . WHO # . might just as WÉLL # be doing the things we're very GÖÖD at # in other sorts of DEPÄRTMENTS # and this is leading me gently to the nation [?] of – – the possibility that a departmental view of an ARTS faculty # (LL S.3.4,863,864-9)

The italicized cleft in (59) exhibits ellipsis of the relative clause, recoverable from the cleft which precedes it.

- (59) Popular etymology shows, in fact the operation of a widespread and powerful linguistic process, analogy. We learn, recollect, and become adept at using language by analogy, that is by

recalling likenesses of meaning, grammatical context, form or sound. We know that cool, coolness, and even cold, are related to each other. It is not surprising, therefore, that our ancestors, knowing that oecern (modern acorn) referred to the fruit of the ac 'oaktree', should assume a connection between the two and believe that -cern should be changed to -corn. In fact, the word oecern is related to oecer 'a field' (modern acre, which has, however, become specialised in meaning), and originally referred to the produce of the fields in general. It is not the observation of likenesses which is at fault in popular etymology, it is the fact that conclusions about the relationships of words, drawn from comparisons, happen to be erroneous. (LOB G5t,58-9,59-61)

(b) Contrastive-given

Throughout the extract in (60) there is a process of identification in operation, as the speakers debate which university body is responsible for handling the matter under discussion. It is almost inevitable that cleft sentences should be used, to facilitate the formulations of contrast as alternatives are proposed. The first cleft (tone unit 1014) belongs in Subcategory 1d and is discussed below. The second is parallel to the first, a revision of it, which counters Speaker B's initial suggestion that it would be 'the Board of Sindies' that would handle the matter. The content of the relative clause, *that'll be putting him forward*, is directly recoverable, the function of the nucleus on *forward* being simply to reinforce A's earlier (almost) identical words.

- (60) B ((I suppose it would)) go to the board of STUDIES
 { WOULDNT it # } # ((REALLY))
 A NÔ # . boards of studies don't . don't DEAL with
 RECOGNITION # ((this is a - { bloody}))
 COMPLICATED { UNIVERSITY # [ə] it's . it's [di ð]
faculty in the SCHOOL
 B oh no they DONT { of COURSE # } # DO they # . NÔ #
 no they DONT # - ((NÔ#))
 > A - { that that puts you FORWARD # } # you SEE #
 B [m] #
 > A . so that it's the faculty of ARTS # . [ə] or the faculty of
 ECONOMICS # or BOTH # within the N F Ô # - that'll be
 putting him FORWARD # (LL S.1.2 1014,1022-6)

In (61) the highlighted prepositional phrase of the cleft contains a contrastively new demonstrative (*this*) in the complement. The relative clause is given: not only is its proposition easy to accept, in an article that is written with the intention of alerting readers' attention to matters with which they should be mainly concerned, but also it picks up words such as *important* and *desirable* earlier in the text. The unmarkedness of the cleft in (61) is not unexpected. Being text-lial it is appropriate that it should not 'draw attention to itself'. Not only is the combination 'contrastive + given' unmarked, but it enables the writer to formulate a sentence with low communicative dynamism and a predominantly textual function.

(61) Danger-spots

But the problem of absorbing immigrants harmoniously into British society is as important to the immigrants as to the British. One of the important conclusions reached by Mr James Wickenden in his valuable study on 'Colour in Britain' is that the danger appears to lie 'where a concentration of immigrants has formed too quickly for an area's capacity to absorb them. Where this occurs there has been violence and the danger of violence and hostility will always be present. As a short term measure it is therefore surely desirable to keep the number of immigrants to a level which can be absorbed.'

The 'open door' policy is of value only so long as genuine hospitality and security can be offered to the newcomers. *It is with this aspect that we should be mainly concerned.* (LOB B15,146-7)

(c) New-inferable

The relative clause of each of the clefts in (62) is realized as a separate tone unit with unmarked nucleus-placement, but in both cases the content of the clause is recoverable by inference (that 'we objected to something', 'we despised something', 'something made the line' are inferable from the description of the antipathy between the school children of Barnsley and Banbury).

- (62) b [ə.m] we were to a certain extent SNOBS I think #
 a [ni] #
 b [ə:] . BECAUSE # . of the BEHAVIOUR # of the people
 who came from BARNSELY # . not . not in the [li: ʔu:] . 1

suppose in a WAY # - [a: æ?] . MY schooling was STARTED # , with . the BAMBURY lot #

(. . .)

- b I suppose we had plenty of MONEY # and we all lived in the Temperton Park ROAD # - , and it was not - - - the difference . perhaps in our rates and RENTALS { compared with the BARNSELYTES # } # that we OBJECTED to # . but they used to have their bus from BARNSELY # which I've PUT in my MEMORIES #

a [m] # . YES #

- > b with the { DOORS } swinging open BEHIND # and shrieks and GIGGLES # and it was . the lack of DECORUM # -

a (- - - laughs) . ((2 sylls))

- > b which we DESPISED so # . [a:] so that to to that EXTENT # we were SNÖBBISH # (LL S.6.4,793-4,801-3)

The cleft in (63) participates in the working out of a contrast: 'it was not so much the weak defence of the Royal Navy team, but the experience and vigour of Swansea . . .'. While the highlighted element is clearly freshly new, the relative clause is inferrable, sharing a semantic relationship with *were too strong for, concede, proved too much for and lost*.

- (63) Swansea were too strong for the Royal Navy on a muddy St Helens pitch in heavy rain yesterday. Having to concede weight forward proved too much for the seamen and they lost by two goals and two tries to a try.

A lack of determined defence in midfield and casual defensive covering allowed Swansea to score tries, but it was the greater experience and vigour of Swansea, with five internationals, which carried the day. (LOB A08,98-100)

(d) Contrastive-inferrable

- 2 The first cleft in extract (60) above is encoded as a single tone unit with compound nucleus. The first peak of information falls on the constituent *the faculty in the school*, which Speaker A opposes to Speaker B's *bourd of studies*, and whose contrastive flavour is reinforced by its non-finality in the tone-unit. The relative clause, although it carries a second peak of information, is indirectly recoverable (*pus forward* could hardly be highly newsworthy, insofar as the

rising-falling nucleus on *school* has already exhausted much of the tone unit's 'quantum' of information potential).

In (64) the cleft would, on a natural reading, be intoned as a single tone unit with the focus (or with the main peak of prominence, if read as with compound nucleus) on *man*. The highlighted noun phrase, *the man*, is thus contrastively new, being contrasted with 'the woman'. The proposition expressed by the relative clause is readily deducible from the fact that, if a dancer fails to execute a turn properly, then he or she is 'at fault'.

- (64) Inexperienced dancers often have difficulty in ending the Natural Turn in the correct alignment; facing centre instead of diagonally to centre, and on the Reverse Turn they will often end facing wall instead of diagonally to wall. *It is usually the man who is at fault*, and the fault is, failure to turn the hips sufficiently between steps 5-6. (LOB E13,160-1)

(ii) Type 2 clefts

The prototypical Type 2 cleft has a short highlighted section which, whether directly or indirectly recoverable, 'sets the context' for the presentation of the new information in the relative clause.

(a) Given-new

In (65) the highlighted element, *Norma Hurley*, though not mentioned previously in the text, is presented as given and thus presumably represents shared background knowledge for both Speaker A and Speaker B. The relative clause, while it conveys a quality of settled non-controversiality, is nevertheless presented as freshly new by the speaker.

- (65) B YEAH # -

A one thing I DIDN'T know # when she was ASKING [a:] AGAIN # it was Norma Hurley that did most of the TALKING # cos she's been . she and Mark Toomy have made most of the ENQUIRIES # but [a:] Mark obviously briefed her to ask [di:] ((relevant)) QUESTIONS # (LL S.3.2,1035)

The cleft sentence italicized in (66) is a prototypical marked cleft. It has a short, deictic, highlighted element, *hereabouts*, which 'sets the context' for the presentation of the 'news' of the sentence.

- (66) Eve Perick
The new films
La Lollo and the hockey girl bully-off

I am happy to report that I saw something this week I have never before witnessed, either in pictures or outside – a budgie-igar playing a drink scene, and playing it with perfect timing and technique.

(...)

Mr Hudson and lady love Lollo find themselves playing charperon (Brenda de Banzie, the official one, has a broken leg) to the girls, who have just been joined by a Jeep-load of boys.

It's hereabouts that the budgie takes to the bottle, but I don't think it was through boredom. (LOB CO6 142)

(b) Given-contrastive

In (67) the highlighted element of the cleft, *us*, is partly situationally given (the set of referents includes Speaker A and his addressee) and partly inferrable ('those who, like us, have interests in grammar'). The nucleus on *them* is contrastive ('the mathematicians, and not the grammarians'). The falling-rising pitch of the nucleus not only reinforces the sense of contrast with earlier material, but also anticipates the addition *not vice versa*.

- (67) C ((I mean)) every transformation word that I've HĒARD #
B YĒS #
C is in AT the moment # in [ði:] course for MATHĒMÁTICS #
B that's RĪGHT # well that's where it all CÔMES from #
C YĒS # and it's { SŌ } . FĀSCINATING # to see the
ANĀLOGY # – and it's much better in the , mathematics
than it is in GRAMMAR I think #
B but it's us that lifted it from THĒM # not
C YĒS #
B vice VĒRSA # (I.L.S.2.5,625)

(c) Inferrable-new

The cleft in (68) is sufficiently short in length to warrant a claim that it would most probably be realized as a single tone unit in reading

aloud. The inferrability of the highlighted element (*influence* being deducible from the reference to 'producing winners') suggests that this tone unit is intonationally unmarked, with nucleus located on the final lexical item *predominates*.

- (68) It is usual for a mare who has produced one winning jumper to produce others, even if not of the same standard, and generally in steeplechasers it is *the influence of the mare which predominates*. (LOB A23,148)

The highlighted prepositional phrase of the cleft in (69), *out of the world of the Greek dispersal*, is implicit in the prior reference to 'the diffusion of Greek language and culture'. The relative clause, by contrast, clearly represents information with a low degree of textual predictability.

- (69) Nearly the whole of Cavafy's life was spent in Alexandria. This, as can be seen when one knows Cavafy, was a fitting background. It was the city founded by Alexander the Great, the city where he was buried, the city above all symbolical of the diffusion of Greek language and culture from the Indus to the far west. Of other Greek cities only Athens and Constantinople have equally powerful associations, and the worlds of Alexandria and Constantinople are, of course, utterly different from the world of fifth-century Athens. *It was out of the world of the Greek dispersal that Cavafy created his personal mythology – a world both of triumph and disaster, a world of courage, of humour and of irony.* Cavafy was the first modern Greek poet who contrived to be patriotic without being romantic. (LOB C12,68–71)

(d) Inferrable-contrastive

In (70) the highlighted prepositional phrase, *through Wesley's efforts*, is implicit in the description of Wesley's activities immediately preceding (his dealings with Joseph Green and the Rector, and his services at Shipton). The relative clause is of low communicative dynamism, with *Green* and *Lincoln* directly recoverable, and *came* likely to be read as focally marked.

- (70) In June, 1730, he noted proudly that he had his 'first pupil', in all probability Joseph Green, the Bible clerk whom he had

introduced to the Rector on 10th June and whom he took to be matriculated two years later. Green's father lived at Shipton, where Wesley often took the service for this friend, the former Lincoln undergraduate, Joseph Goodwin. *It was probably through Wesley's efforts that Green came to Lincoln.* He was soon calling on Wesley, who lived in rooms just above him in College, at ten every morning, presumably for tuition. (LOB G18,43-4)

(iii) *Type 3 clefts*

Type 3 clefts typically have a circumstantial adjunct or adjuncts as highlighted element, and the bulk of the propositional content in the relative clause. With their capacity to imbue information, even though freshly communicative, with a character of non-controversiality they give the impression that the listener/reader is simply being 'put in the picture', or 'brought up to date' with information to which others will already be privy. (This impression has several sources: the dependency, and presupposedness, of the relative clause; and the regular association of the highlighted element with contrastiveness, which is never a totally fresh form of information.)

The claim that Type 3 clefts contain predominantly new information is supported by their distributional behaviour (and note that this behaviour is seen as a reflection of the textual structure of the construction, rather than being used to define it). Type 3 clefts either occur discourse-initially or stage-initially, or have the potential to be so distributed (those containing contrastive information being limited to stage-initial position, actual or potential). They thus serve to moderate the brusqueness which might result, in the corresponding non-cleft, from the presentation of unmitigated new content in topic-sentence position.

(a) *New-new*

The cleft in (71) occurs, as do many clefts of this subtype, in the register of press reportage. As is typically the case with the subtype the highlighted element – a circumstantial adjunct – is, though informationally new, clearly less communicatively dynamic than material in the relative clause.

(71) *THE GRANDSTAND ELITE PLAY ONE-UPMANSHIP*

It was 6.55 yesterday morning when the Duke of Norfolk wandered across the green lawns of Ascot. The new £1,000,000 stand he had watched rise, brick-by-brick, was clouded with rain. And the wind had littered the paddock with leaves.

His Grace ordered the leaves to be removed and the sweepers came out. (LOB A10,5-6)

Extract (72), from the *Sunday Express*, contains an example of the potentially-discourse-initial Type 3 cleft that is common in newspapers. Neither the horse referred to in the cleft, nor its trainer, nor the 'racecourse debut' scheduled for December, have been previously mentioned.

(72) In Northumberland, the sparkling Kerstin's full brother and sister – Vindicated (now with Guy Cunard) and Lady Nenagh – made their mark for Verly Bewicke, many of whose horses are related to each other.

Another North-country star, Rough Tweed, winner of Manchester's Champion Novice 'Chase last April, will have a full-brother running over here this season.

It is four-year-old Holy Loch, trained by Bobby Norris in Northamptonshire, who makes his racecourse debut – over hurdles – early in December. (LOB A23,165-7)

A clear example of a Type 3 cleft is that in (73), with the characteristic scene-setting adverbial as highlighted element. The presentation of freshly new material in the relative clause (where, in the agnate non-cleft, it would not be syntactically subordinate) has the rhetorical effect of imbuing it with a non-controversial character. As a result the reader is made to feel that s/he is being made privy to a generally known piece of information.

(73) *A Fastidious Composer*
'Journal' debut at Cheltenham
 From Donald Mitchell
 Cheltenham, Thursday.

It was not long ago that Richard Rodney Bennett composed a 'Calendar' for chamber ensemble. Now he has written a 'Journal' for orchestra which was given its first performance in the Town Hall, Cheltenham, to-night by the B.B.C. Symphony Orchestra conducted by Norman Del Mar. (LOB C01,73-4)

(b) New-contrastive

In (74) the relative clause of the cleft is realized as a tone unit with a compound nucleus. The first peak of prominence falls on *we*, a non-lexical item. The second peak, meanwhile, falls on an inferrable item, *concerned* (not only is Speaker B expected to infer a link between *not irrelevant* and *concerned*, but also to accept that the referent of *we* is, in the context of the present discussion, likely to 'be concerned about' certain matters). The highlighted element, which spans two tone units, participates in the formulation of a contrast, but is clearly not contrastive (in the sense of 'given information newly interpreted').

- (74) A NO # -- [ə:m --- wə ə:] . we've got to [ə:] --- DECIDE # -- what the structure of faculty BOARDS # and boards of STUDIES # . ARE # -- going to be in the FUTURE #
 B is THAT the actual { terms of REFERENCE # } #]
 thought that ((it)) was more [ə] you know [ə] ((if if you if you want to have)) --
 A ((2to3 sylls the committee's)) not in EXISTENCE YET #
 B ((YES #)) . if you want to have PHILOSOPHY # , and [ʔə]
 . MATHEMATICS # ((as your your) two possible
 SUBJECTS # as an UNDERGRADUATE # then you can
 DO those # .
 A (OH) NO #
 B ((then)) this isn't your -- ((CHORE)) # --
 A NO # , it's not it's not IRRELEVANT to it # -- [ə:m] . but --
 -- it's the -- academic STRUCTURE # .
 B ((2to3 sylls))
 > A of . the UNIVERSITY # that . that WE'RE CONCERNED
 about # (LL S.1.2,1336,1338-9)

(c) Contrastive-new

The cleft in (75) displays marked nucleus-placement in the first tone unit realizing the superordinate clause. the contrastiveness of the demonstrative, *these*, being reinforced by the choice of falling-rising nucleus. Speaker A reserves the 'news' of his desire to begin a Ph.D for the relative clause of the cleft.

- (75) A there's ANOTHER interesting # SIDE to this # and that IS
 that # . EVERYONE # is very BUSY NOWADAYS # .

TRANSFERRING written language # from ONE written
 medium # to ANOTHER # . typewriting to PRINT #

d [m]

A HAND to print # . blackboard to . HANDWRITING # --
 you you get different . CONVENTIONS for layout # differ-
 ent conventions for PUNCTUATION # -- different OTHER
 conventions # for { ALL these different } MEDIA # -- and
 INCIDENTALLY # this is important for NOTETAKING
 # -- and it's somewhere along THESE lines # that I would
 like to .

d [m]

> A do a PH D # -- (LL S.2.4,1367-8)

Within the context of the argument in (76) that non-swimmers are more likely to drown than swimmers, *the swimmer* in the highlighted section of the cleft is given, but newly identified as the relevant member of the set of two.

- (76) Letters to the Editor
 Subways preferred to baths

Sir, -- Your correspondent S. Armitage quotes a figure of nearly 5,000 people drowned in and around Britain in 1960. I know nothing of the accuracy of these figures, although Saturday's 'Echo' mentions a figure of 4,000 every year.

The point I wish to make is that all these casualties did not occur among the non-swimming members of our population. In fact I venture to suggest it is probable that the greater proportion of these unfortunate people could swim and in fact might not have been drowned had they been non-swimmers.

It is so often the swimmer that ventures out, gets into difficulties and is rescued, if there is time. Non-swimmers are content to paddle, sun-bathe and splash about generally at the seaside and very rarely I think enter rivers. (LOB B24,13-14)

(d) Contrastive-contrastive

In (77) both *David* and *Ingrid* are presented intonationally as contrastively new. Both have referents known to the interlocutors, but the marked nucleus in both cases (reinforced in the case of *David* by the

choice of falling-rising nucleus) indicates that they are being 'newly identified'.

- (77) c YĒAH # how's DŌN's jobs going # [ɔ:m] perfectly WĒLL I
GATHER # ((1 syll))
h because David knew DŌN # you KNĒW #
A oh YES # I KNOW # I KNŌW #
c (laughs)
b (laughs)
A it was [ɔ:] {THRŌUGH }; DĀVID that [ɔ:m] # ĪNGRĪD
(fmet Don # [ɔ:])) (LL S.4.4, t48-9)

The cleft in (78) reverses the relationship between the 'young' and 'old' presented in the preceding sentence. Whereas there it was the old who passed on something (memories) to the young, in the cleft it is the young who pass on something (recipes) to the old. Thus both the highlighted element and the complement of *from* in the relative clause have a contrastive interpretation ('mother, not daughter'; 'her daughter, not mother').

(78) The Full Table

Happy, happy families! Never before have Britain's larders been so well stocked. Supplies of meat and dairy produce were substantially higher last year than in 1959.

Lucky, lucky housewives! To have such a splendid variety of goods to choose from.

Not so long ago older folk were reminding young wives, harassed by shortages, of the good old days of abundance. *Now it is mother who picks up recipes from her daughter.*

The dinner table is the best answer to the grumblers in Britain today! (LOB B04,31-2)

6.3.3 Clefts and theme-rheme structure

In Section 5.2.3 above reference was made to two approaches to the thematic analysis of cleft constructions that have been suggested by Halliday. The difference between Halliday's earlier analysis, in which the predicative *it* is taken to be part of the rheme, and his later 'metaphorical' analysis in which the superordinate clause is all thematic, is regarded as relatively insubstantial for the purposes of the

present discussion. Of primary interest here is the choice of highlighted element (neither the *it* nor the copula carries any appreciable degree of 'thematic dynamism', given their almost obligatory initial position: occasionally clefts display thematic fronting, with the highlighted element supplanting the *it* as a marked theme). In the discussion that follows, therefore, the theme of a cleft will be understood to be precisely its highlighted element. The thematic status of the *it* and copula will not occupy any further attention.

Like reversed pseudo-clefts, but unlike basic pseudo-clefts, cleft constructions display a structural mapping of the theme on to the highlighted element/identifier. Nevertheless (as argued in Section 5.2.3) clefts and reversed pseudo-clefts achieve thematic prominence in different ways. In reversed (and basic) pseudo-clefts the theme represents one of two members participating in an equative relationship, and carries a prominence that is essentially ideational in flavour. In clefts the elements realizing the theme and rheme are not reversible, and the emphasis falls less upon the identity between them than upon the predication of one part in the structure, the theme. The theme in clefts thus carries a 'textual' form of prominence.

There follow examples from LL and LOB which illustrate the textual thematic prominence encoded by the cleft construction. This prominence is most evident in clefts with a non-ideational highlighted element. Such clefts rarely have an acceptable pseudo-cleft counterpart, a fact which follows from the stronger expectation than arises with the cleft that the theme of the reversed pseudo-cleft will be nominal. Consider (79):

- (79) But there were many problems that exercised his active mind. 'I was filled with questions', he says, 'and I have to carry these questions about with me unanswered. *It was thus that I reached my eighth year.*' (LOB G32,42)

Here the conjunctive adjunct which serves as predicated theme (*thus*) is almost obligatorily thematic, and is thus an appropriate candidate for the thematic highlighting conferred by the cleft construction. Although *thus* would be thematic in the corresponding reversed pseudo-cleft, the reluctance of the pseudo-cleft to accept a non-nominal highlighted element robs it of complete acceptability ('?Thus was how/the way I reached my eighth year'). The corresponding basic pseudo-cleft ('?How I reached my eighth year was *thus*') is even less acceptable, as a result of the relegation of *thus* to rhematic status. (The way in which reversed pseudo-clefts satisfy the preference of the

construction for nominal themes is by selecting as theme either demonstrative *this* or *that*, these being interpretable as the pro-noun modifier of the corresponding prepositional phrase, in *this/that* way: 'this/that was how I reached my eighth year'.)

The italicized cleft in (80) presents as theme a prepositional phrase with an explicitly text-referential function. The combination of this function along with the fact that, as a prepositional phrase, it does not lend itself readily to highlighting in an identifying relationship (as in pseudo-clefts), renders the cleft here a particularly suitable candidate for the textual prominence that the cleft construction generates.

- (80) For the true native painter, on the other hand, there is no margin between his intention and his result: he paints to the exact limit of his vision. It is exactly in his humble capacity to be satisfied with this that his naivete or lack of sophistication lies. *It is exactly in this that his appeal lies.* (LOB G48,50)

An option that is available with clefts, but not reversed pseudo-clefts, is the possibility of marked theme. Throughout most of the extract in (81) the method of development is a series of references to Mr Gilliatt. The marked thematic cleft explicitly asserts the thematic status of *he* (Mr Gilliatt), avoiding the possibility of even the temporary ambiguity that might result were it to be decoded as a referential (and therefore normally thematic) pronoun.

- (81) # OPPOSITE THIS # is DUFFIELD House # - Duffield HOUSE # . was the HOME # . for POSSIBLY # two GENERATIONS # of the GILLIATT . family # - Mr GILLIATT # . AS # . a young BOY # was the CHURCH warden at Stoke Poges CHURCH # and he HAD been Stoke . [a:] WARDEN # for about twenty-five YEARS # - *he* *it* WAS # *who* built . Saint Paul's CHURCH # (in Stoke ROAD) . HIMSELF # at his own EXPENSE # - so he was a great BENEFACTOR # (LL S.12.6.1204-7)

Clefts with *yes/no* interrogative superordinate clause, with negated superordinate clause, or with a combination of these (as in (82)), display more clearly than their mood/polarity-unmarked counterparts the textual form of prominence associated with clefts. In such cases it is the identity of the theme that is questioned, disputed, or both, rather than the ideational content of the message. Conversion of the cleft in (82) into its reversed pseudo-cleft counterpart would alter the

sense of the theme. Compared with the textually prominent theme of the cleft ('Is it not the case that my theme is to be identified as *the existing organism*, and only that?'), the theme of the reversed pseudo-cleft version ('Is not the existing organism what determines the use and quality of the new material introduced?') both questions and disputes the role of the 'existing organism' in the 'determining' process.

- (82) Do they pick up nothing from their predecessors? If they do not, how is it that the same flour and mutton eaten at the same table will turn to four different conditions as regards colour and organic quality, in four different persons? *Is it not the existing organism that determines the use and quality of the new material introduced?* and how could this be, except on the principle of transmission of quality? (LOB D14,64-5)

The textual thematic prominence of the cleft construction is also displayed clearly in *wh*-interrogative clefts. *Wh*-interrogative clefts, as in (83), more consistently retain the word order of the corresponding non-cleft than do their declarative counterparts, given that in the non-cleft, as in the cleft, the *wh*-item will be thematic. The speaker/writer can then encode the extra semantic features that are expressed by the cleft sentence (exclusiveness, textually prominent thematicity, and so forth) simply by interpolating '*it he* + relative' after the *wh*-item, and adjusting the verb group accordingly.

- (83) Otherwise it might not easily be recognised that one's true ambitions lie 'across the river', and that Local Government is merely a lay-by, on the road to Westminster.

If it is suspected that this is not the case, then one is hastily dismissed as a crank, and an oddity - and ignored from then onwards.

Why? *What is it that makes people look upon Local Government as dull, unexciting, and unrewarding?* And the people involved, as failures in the 'Grand National' Stakes - or just non-starters? (LOB F16,166-7)

As noted in Section 4.2 above, a very popular type of cleft is that with 'zero-theme' (where 'zero' is to be interpreted in an ideational sense only). Through the cleft construction, strongly thematic status is given to elements of modality, polarity, tense, aspect, or combinations of these. It is the rarity with which these elements are

thematic in any other construction that explains the strongly foregrounded character that they exhibit in the cleft construction. The only way in which such elements could be highlighted in a non-cleft construction would be via an adjunct (compare 'It may be that Tom is unwell' and 'Maybe Tom is unwell'). In (84) it is an indication of (epistemic) modality that is foregrounded.

- (84) # the safe CONTRACEPTIVE # -[ə: ə: a: i i i] is ENTIRELY # [ə sə ə ə] STANDING # . the world on its HEAD # . [æ] and it may BE # . that the FAMILY # . will DISAPPEAR. { as the BASIS } of CIVILIZATION # } # } (LLS 6.4, 656-8)

Finally, in keeping with the textual thematic prominence expressed by the construction, clefts occur with a wide range of thematic prepositional phrases. In (85) the textual sense of the theme is reinforced by the anaphoric demonstrative *that*, and in (86) by the negation of the superordinate clause.

- (85) b (. clears throat) YES but # . there's a tendency then for her to go and repeat EVERYTHING # [tu əs] - - - ANNOYANCE nf
a WHY d'you say ((that)) #
> b everybody # - - -
a [u] it was for that REASON # that I thought of DOING this # (LLS 4.1, 206-7)

- (86) The position is that, whereas this claim of 'objectivity' is of extreme interest as evidence of Schiller's aesthetic consciousness and of his efforts to bring it to terms with his theoretical reflections, he does not in fact substantiate the claim in its more far-reaching implications. Besides, it is not in this area that the main advance is made. (LOB 153, 44-5)

In conclusion, a brief examination of several examples will serve to confirm further the claims advanced in this section. In particular, I have argued that the theme of the cleft construction is semantically distinctive, and that the cleft construction is very 'flexible' thematically, accepting as the theme constituent types not favoured by the pseudo-cleft construction. It is perhaps as much on account of their thematic flexibility, as it is their unambiguous indication of the informationally focal constituent, that clefts are so popular - relative to pseudo-clefts - in written discourse. The tentative suggestion has already been made in Section 5.3.5 above that while information

structure is the primary form of textual organization in speech, this role is served in writing by theme-rheme structure. Not only is the cleft construction thematically flexible, but also the unambiguous mapping of theme on to new information in the unmarked instance (for which there is a clear preference in writing), gives it a particularly useful function in the textual organization of written discourse.

The first extract is from LL and contains three clefts:

- (87) > A (. .) # - [ə:m] you're very kind old SAM # - - BLÉSS you # well that finishes THÁT # . [ə:m] ((now)) what was the OTHER thing { I wanted to ASK you # } # . [i] is . is it this YÉAR # that [ə:] NÍGHTINGALE goes # - -
B [ə:] no NÉXT year # - -
A [ə:m] . sixty - [f] - four SIXTY-FIVE .
B SIXTY-FIVE # YÉAH #
> A ((I thought it was BÉFÔRE)) sixty-five # so it's not until - next year that the job ((will be)) ADVERTISED #
B January I suppose there may be an interview round about JANUARY # (LL S.1.1, 238-9, 244, 245)

Speaker A's question, at the close of his first turn in the extract, is appropriately encoded in cleft form, since it seeks clarification on the identity of one member of a set of possible candidates (that is, the year of Nightingale's departure which, as the ensuing discussion indicates, the interlocutors are having some difficulty in distinguishing from the years that Nightingale does not depart). The cleft has two possible reversed pseudo-cleft versions, one with a temporal relative ('is this year when Nightingale goes?') and the other with a nominal relative ('is this year the one (that) Nightingale goes?'). The latter seems to be the more acceptable, probably because the choice of relative generates the impression that the antecedent (*this year*) is nominal in function (as well as form), rather than an adjunct. While there are subtle differences in the nature of the exclusiveness conveyed by the original cleft (textual: 'this year - and no other year - is my topic') and the reversed pseudo-cleft(s) (ideational: 'this year - and no other year - is the one that Nightingale goes'), the feature of exclusiveness is lost altogether in the non-cleft version. Whereas the cleft questions the existence of a relationship of exclusive identity between *this year* and 'Nightingale's time of departure' (the answer is either 'yes it is' or 'no it isn't'), the non-cleft - 'does Nightingale go this year?' - questions the validity of a proposition concerning a (future) event.

The clause *it was before sixty-five* has been accepted as a cleft, with ellipsis of the relative clause *that Nightingale goes*, on the basis of the parallel that is established, on this reading, with the cleft in Speaker A's first turn. The alternative possibility, here rejected, is that it is referential (to 'the time that Nightingale goes') and the clause of which it is the subject is non-cleft. With its prepositional phrase-theme the cleft does not convert readily into a pseudo-cleft ('before sixty-five is when Nightingale goes'), but a non-cleft version is available ('Nightingale goes before sixty-five'). The agnate non-cleft with fronted prepositional phrase ('before sixty-five Nightingale goes') would be thematically cohesive, and yet less appropriate than the original as a consequence of the absence of any exclusiveness feature.

In (88) a series of clefts is used to structure the argument thematically around the notion of the 'domain' or 'area' in which Schiller's primary message is grounded: 'in the issue of the "objective" principle: 'in the vindication of beauty as a function of the human totality'; 'in aesthetic theory proper'. The cleft construction is ideally suited to the task of representing a series of adverbial constituents as unmarked themes, and moreover associating these with a feature of exclusiveness that reflects the author's attempt to identify the relevant area in which Schiller's philosophical contribution lies.

- (88) (. . .) Schiller does not meet directly Kant's main argument for this view, which is, in essence, that the aesthetic judgement rests on a subjective pleasure, which cannot itself be the product of a deduction. Indeed there is already evidence that it is *not* in this teasing issue of the 'objective' principle, in the sense developed above, that Schiller's vital concern lies, but rather in the vindication of beauty as a function of the human totality. Schiller is dissatisfied with Kant's manner of excluding rational form entirely from the province of beauty. He concedes the necessity of a sharp distinction between perfection logically apprehended, and the beautiful, but considers that Kant's solution is misguided and impoverishes the idea of beauty:

(Long quotation)

It is here that Schiller's more valid challenge lies, from the beginning of his 'Auseinandersetzung mit Kant'. Here, that is, in aesthetic theory proper. (LOB J53.113-15, 115-16, 122-3, 123-4)

The adjunct themes of the two clefts in (90) would, if preserved as themes in the corresponding non-clefts, not be malcoherent, but

would be arguably less apt than the original clefts. This is because the inherently anaphoric deictic items, *this* and *then*, have a predominantly textual function and thus lend themselves readily to the textual form of highlighting associated with the theme of the cleft sentence construction.

- (90) (. . .) At last, however, the nations of the earth will become so decimated by war, famine and self-destruction that Zechariah speaks of 'every one THAT IS LEFT of all the nations which come against Jerusalem' (Zech 14:16): *It is thus that Israel, as the vehicle of the true seed on earth, come into their own, for then 'Israel shall blossom, and bud, and fill the face of the world with fruit' (Isa. 27:6); it is then that they 'enlarge the place of their tent' and their seed 'shall inherit the Gentiles' (Isa. 54:3) even as their fathers in small yet typical measure 'inherited' the land held by the Amorite (Deut. 2:31).* (LOB D12.83-6, 86-9)

The cleft construction thus serves an important and distinctive role in the textual organization of discourse. The mapping of theme and identifier and, in unmarked instances, of new information, gives rise to an exclusiveness feature. This, in combination with the structural feature of predication, produces a distinctive textual form of thematic prominence. Often, and especially when a non-nominal theme is involved, no acceptable pseudo-cleft alternative is available for a cleft, while substitution of the corresponding non-cleft may result in a reduced degree of cohesiveness, as a result of the loss of particular semantic features.

7 Clefts, pseudo-clefts, and register variation

This chapter examines the distribution of cleft and pseudo-cleft constructions in LOB and LL, and across the various registers represented in them. Particular attention is paid to the differences between spoken and written language.

7.1 CLEFTS AND PSEUDO-CLEFTS IN SPEECH AND WRITING

In this section I shall compare the distribution of cleft and pseudo-cleft constructions in LL and LOB, taking the two corpora as broadly representative of speech and writing respectively. Figures for the two corpora are presented in Table 7.1, in which frequencies are given for the fifteen genres of LOB, along with subtotals for the 'informative' and 'imaginative' subgroups, and Table 7.2, in which the 87 texts of LL are subclassified – as in Table 2.2 – according to a variety of register features, including mode of communication (dialogue or monologue), tenor relationships between participants (intimates, equals, or disparates), nature of speech situation (private or public), and method of recording (surreptitious or non-surreptitious).

The first thing that strikes one is the quite uneven distribution of clefts and pseudo-clefts in speech and writing. Pseudo-clefts greatly outnumber clefts in speech (by a ratio of 623:187, or 3.3:1), while clefts outnumber pseudo-clefts in writing (by a ratio of 565:410, or 1.3:1). In Table 7.3 the two constructions are ranked according to their popularity in speech and writing and, in addition, the relative popularity of basic and reversed pseudo-clefts in the two modes is given.¹ As can be seen, reversed pseudo-clefts outnumber basic in both speech and writing.

Table 7.1 Frequencies of pseudo-clefts and clefts in LOB*

Category	No. of texts	Pseudo-clefts			Clefts	Total
		Basic	Reversed	Total		
A Press: reportage	44	1.0 (9)	1.7 (15)	2.7 (24)	2.7 (24)	5.5 (48)
B Press: editorial	27	4.6 (25)	3.1 (17)	7.7 (42)	7.0 (38)	14.8 (80)
C Press: reviews	17	1.5 (5)	1.5 (5)	2.9 (10)	7.6 (26)	10.6 (36)
D Religion	17	2.4 (8)	2.9 (10)	5.3 (18)	6.8 (23)	12.1 (41)
E Skills, trades and hobbies	38	2.0 (15)	1.1 (8)	3.0 (23)	5.1 (39)	8.2 (62)
F Popular lore	44	2.2 (19)	1.6 (14)	3.8 (33)	6.3 (55)	10.0 (88)
G Belles lettres: biography, essays	77	2.1 (33)	1.6 (24)	3.7 (57)	6.8 (104)	10.5 (161)
H Misc. (Govt docs, reports, catalogue, etc.)	30	1.5 (9)	0.5 (3)	2.0 (12)	1.8 (11)	3.8 (23)
J Learned and scientific writings	80	1.1 (17)	0.8 (13)	1.9 (30)	5.1 (82)	7.0 (112)
Total for informative prose	374	1.9 (140)	1.5 (109)	3.3 (249)	5.4 (402)	8.7 (651)
K General fiction	29	2.1 (12)	2.9 (17)	5.0 (29)	7.4 (43)	12.4 (72)
L Mystery and detective fiction	24	2.9 (14)	5.0 (24)	7.9 (38)	6.7 (32)	14.6 (70)
M Science fiction	6	1.7 (12)	5.0 (6)	6.7 (8)	2.5 (3)	9.2 (11)
N Adventure and Western fiction	29	1.0 (6)	4.0 (23)	5.0 (29)	5.5 (32)	10.5 (61)
P Romance and love story	29	1.7 (10)	5.7 (33)	7.4 (43)	7.4 (43)	14.8 (86)
R Humour	9	4.4 (8)	3.3 (6)	7.7 (14)	5.5 (10)	13.3 (24)
Total for imaginative prose	126	2.1 (52)	4.3 (109)	6.4 (161)	6.5 (163)	12.9 (324)
Total		1.9 (192)	2.2 (218)	4.1 (410)	5.7 (565)	9.8 (975)

* Frequencies = tokens per 10,000 words; () = raw frequencies

Table 7.2 Frequencies of pseudo-clefts and clefts in LL*

Sub-group	Broad description	Dialogue or monologue	Private or public	No. of texts	Pseudo-clefts		Clefts	Total
					Basic	Reversed		
A	Face-to-face conversation	Dialogue	Private	34	6.4 (108)	10.9 (185)	5.3 (90)	22.5 (383)
B	Face-to-face conversation	Dialogue	Private	12	3.8 (23)	11.3 (68)	3.2 (19)	18.3 (110)
C	Telephone conversation	Dialogue	Private	10	6.0 (30)	4.8 (24)	2.4 (12)	13.2 (66)
D	Discussion, interview, debate	Dialogue	Public	12	11.3 (68)	7.5 (45)	4.3 (26)	23.2 (139)
E	Spontaneous oration: court case, parliament	Dialogue	Public	3	8.7 (13)	4.0 (6)	4.7 (7)	17.3 (26)
F	Spontaneous oration: radio	Monologue	Public	2	4.0 (4)	5.0 (5)	3.0 (3)	12.0 (12)
G	Radio commentary	Monologue	Public	7	1.1 (4)	0.3 (1)	2.9 (10)	4.3 (15)
H	Prepared speeches	Monologue	Public	7	6.3 (22)	4.9 (17)	5.7 (20)	16.9 (59)
					6.3 (272)	8.1 (351)	4.3 (187)	18.6 (810)

* Frequencies = tokens per 10,000 words; () = raw frequencies

Table 7.3 Ranking of pseudo-clefts and clefts in speech and writing

Rank	Construction	Mode	No. of tokens	Frequency (per 10,000 words)
1	pseudo-cleft	speech	623/435000 wds	14.3
2	cleft	writing	565/1000000 wds	5.7
3	cleft	speech	187/435000 wds	4.3
4	pseudo-cleft	writing	410/1000000 wds	4.1
1	reversed pseudo-cleft	speech	351/435000 wds	8.1
2	basic pseudo-cleft	speech	272/435000 wds	6.3
3	reversed pseudo-cleft	writing	218/1000000 wds	2.2
4	basic pseudo-cleft	writing	192/1000000 wds	1.9

Why should it be that pseudo-clefts, and in particular reversed pseudo-clefts, outweigh clefts so markedly in popularity in the spoken corpus? Why should clefts be so relatively popular in the written corpus? The clues are to be found. I shall argue, in the different textual functions of these constructions.

Let me characterize some of these differences in broad terms, before subjecting them to scrutiny through a more detailed analysis of the distribution of clefts and pseudo-clefts across the various registers of LL and LOB. Basic pseudo-clefts, I have argued in previous chapters, are characterized by an explicit specification, and conflation, of structures representing aspects of 'shared knowledge'. They display a consistently close mapping of theme, givenness, and presupposition on to the subject relative clause, which is presented to the addressee as representing information that s/he should be prepared to accept as non-controversially recoverable. Basic pseudo-clefts thus offer the speaker a means of specifying precisely, before the announcement of the 'message', the background knowledge to which the addressee is expected to have access. Within the flow of speech, they function as indices, embodying in the relative clause both a backward-looking component (given/presupposed) and a forward-looking component (theme).

Reversed pseudo-clefts also exhibit structural features which provide them with a special indexing role that is particularly suited to the dynamic, snowballing organization of spoken language.² In reversed pseudo-clefts the theme is typically a text-refereutal demonstrative. The text-anaphoric function of the theme, in conjunction with the exclusive equation of this item with backgrounded material in the relative clause, provides the construction with a special 'internal'

referencing' function. Reversed pseudo-clefts often contain minimal 'news', having an almost cliché quality (as in *that's what happened*, *that's what he said*) which enables them to draw together various threads in a text.

Cleft constructions, by contrast, exhibit properties that render them particularly suitable to written language. In both unmarked and marked Type 3 clefts the theme is associated with new information. As such, they may be used by the writer as a means of directing the reader into a particular interpretation of the information structure (namely one where the theme, which is normally associated with given information, is unmarkedly newsworthy). This function of clefts in the written language is to be interpreted in light of the nature of the latent intonation of writing. In writing, unless there is good reason for an alternative interpretation, the expected intonation will be unmarked (see further Section 5.3.1 above, and Halliday 1987).

Another explanation for the relative popularity of clefts in writing lies in the markedness of the theme/new combination, which tends to be of the contrastive kind. This is perhaps part of the reason for the popularity of clefts in 'opinionative' texts (see further Section 7.3 below): many of them occur in contrastive formulations of the type 'It is not *x* . . . : it is *y* . . .', which are commonly associated with argumentative and persuasive discourse.

One further reason for the comparative popularity of clefts in writing may be their structural similarity to impersonal constructions (such as 'It is said that . . .', 'It is true that . . .'), which imbues them with a depersonalized quality and a formality that is often out of place in informal spoken genres.

As a final observation in this section, it is worth noting that the distributional behaviour that I have noted for clefts is largely substantiated by Biber's studies (as reported above). Biber found clefts to be a positively-loaded feature on Factor 2 (interpreted as 'abstract versus situated content' (and most common in official documents and academic prose, but least common in conversation)).³

7.2 CLEFTS AND PSEUDO-CLEFTS IN SPOKEN REGISTERS

In Table 7.4 figures are presented based on several subgroupings of the text categories of LL (as discussed in Section 2.2 above; see in particular Table 2.2). A comparison of the 'private' and 'public' varieties does not reveal vast differences in the text frequencies for clefts and basic pseudo-clefts. Reversed pseudo-clefts, however, are twice as popular in private texts (9.9 tokens per 10,000 words) as in

public texts (4.8). A comparison of the 'dialogue' and 'monologue' subgroupings reveals that pseudo-clefts are strongly preferred in the former (by a ratio of 2.4:1), but clefts less strongly so (by a ratio of 1.3:1). Once again it is reversed pseudo-clefts that display the most striking variation in distribution (with a ratio of 2.9:1, compared to 1.8:1 for basic pseudo-clefts).

It would appear that pseudo-clefts are, relative to clefts, more popular in spoken discourse characterized broadly speaking by informality, personal interaction, context-dependency and privacy of situation. Bearing in mind the general 3.3:1 ratio of pseudo-clefts to clefts in LL, it may be noted that the proportion of pseudo-clefts climbs higher than the average in private discourse (where the ratio is 3.5:1) and dialogic texts (where it is 3.6:1), but slips lower than the average in public texts (where it is 2.9:1) and monologic texts (where it is 1.9:1).

Tenor relationships between participants appear to influence the occurrence of pseudo-clefts. For texts in which the speakers are intimates or equals, the number of pseudo-clefts is generally close to or above the corpus average of 14.3 per 10,000 words (16.0 in Category S.4, 19.2 in S.5, 14.6 in S.8). The sole exception is S.7, where the text frequency is 8.6. On the other hand, for texts in which the speakers are dispartes and distants (including monologue), the reverse is generally the case (10.6 in S.6, 12.7 in E, 9.0 in F, 1.4 in G, 11.1 in H).

Further evidence for the influence of tenor relationships on the ratio of basic to reversed pseudo-clefts is to be found in a comparison of the three categories of telephone conversation referred to in Section 2.3 above. As we move from S.7 to S.9 (that is, from relatively intimate/informal dialogue to relatively distant/formal dialogue), the basic:reversed ratio changes as follows: in S.7 it is 4:9 (or 0.4:1), in S.8 it is 17:12 (or 1.4:1), and in S.9 it is 9:3 (or 3:1).

Reversed pseudo-clefts are considerably more popular in private/dialogic texts than basic pseudo-clefts. A comparison of the figures for reversed pseudo-clefts in LL Categories A, B, and C (see Table 7.2) reveals that the favoured conversational mode for the construction is face-to-face, with intimates or equals as participants. The frequency of reversed pseudo-clefts is at its highest (11.3 per 10,000 words) in Category B, marginally lower (10.9) in Category A, where some of the conversations include interlocutors of disparate status, and significantly lower (4.8) in Category C, where the channel is telephone rather than face-to-face.

The general pattern of preference for reversed over basic pseudo-

Table 7.4 Frequencies of pseudo-clefts and clefts in subgroups of LL*

Text categories	Broad description	Number of texts	Pseudo-clefts		Clefts	
			Basic	Reversed	Total	Total
ABC	Private conversation	56	5.8 (161)	9.9 (277)	15.6 (438)	4.3 (121)
DEFGH	Public non-conversation (interviews, lectures, commentary, speeches, etc.)	31	7.2 (111)	4.8 (74)	11.9 (185)	4.3 (66)
ABCDE	Dialogue	71	6.8 (242)	8.1 (288)	14.9 (530)	4.3 (154)
FGH	Monologue	16	3.8 (30)	2.9 (23)	6.6 (53)	4.1 (33)
S.7	Telephone conversation between intimates	3	2.7 (4)	6.0 (9)	8.7 (13)	1.3 (2)
S.8	Telephone conversation between equals	4	8.5 (17)	6.0 (12)	14.5 (29)	3.5 (7)
S.9	Telephone conversation between superiors	3	6.0 (9)	2.0 (3)	8.0 (12)	2.0 (3)

* Frequencies = tokens per 10,000 words; () = raw frequencies

clefts in dialogic texts is disrupted in two notable cases. In conversations conducted over the telephone (Category C), where, as noted above, the frequency of reversed pseudo-clefts drops markedly, it is in fact outstripped by that of basic pseudo-clefts. Likewise in public dialogue (Categories D and E) basic pseudo-clefts outnumber reversed by a ratio of 1.6:1. What is common to these two cases is a level of formality (deriving, in the case of Category C, from the non-face-to-face channel of communication, and in Categories D and E from the public nature of the speech situation) which is lacking from the private spontaneous conversations in A and B. In their preference for basic as opposed to reversed pseudo-clefts, Categories C, D and E are similar to the 'public' subgroup, where the basic:reversed ratio is 1.5:1, and the 'monologue' subgroup, where it is 1.3:1.

There would seem to be several register factors which influence the relative frequency of occurrence of basic and reversed pseudo-clefts. Generally speaking the proportion of reversed pseudo-clefts increases with the informality of the situation, and with the degree of intimacy or acquaintance of the interlocutors. Thus reversed pseudo-clefts outnumber basic in face-to-face conversations, all of which are informal in LL, and most of which are between intimates and equals. However basic pseudo-clefts outnumber reversed in monologues, public dialogue, and telephone conversation, in all of which a greater level of formality tends to be introduced by the greater degree of distance – both physical and, in the case of lectures and radio commentary, personal – between speaker and hearer(s).

Finally, it can be said of clefts that they are – relative to pseudo-clefts – more popular in monologue than dialogue. The text category with the highest frequency of clefts (2.9 per 10,000 words) is H (prepared oration: dinner speeches, sermons, lectures, and court cases). It would seem that the frequency of clefts is influenced positively by the formality of the speech situation and the extent to which the speech is premeditated. (Of the eight categories in Tables 2.2 and 7.2, Text Category H is the closest in nature to writing and, as we have already seen, clefts outnumber pseudo-clefts in writing.)

The influence of tenor relationships on the frequency of occurrence of clefts is difficult to assess, given their comparatively small numbers. It may be worth noting, however, that the two text subcategories for which the frequency of clefts is significantly lower than average (1.6 per 10,000 words in S.5.8–11, and 1.4 in S.7) occur within the area of conversations among people having equal and intimate status.

7.3 CLEFTS AND PSEUDO-CLEFTS IN WRITTEN REGISTERS

While the overall ratio of clefts to pseudo-clefts in LOB is 565:410, or 1.4:1 (see Table 7.1), in the 'Informative Prose' categories clefts are comparatively even more popular (the ratio being 402:249, or 1.6:1), but in the 'Imaginative Prose' categories their popularity slipped (with the numbers of clefts and pseudo-clefts being almost identical). Whereas basic pseudo-clefts are marginally more popular than reversed pseudo-clefts in informative prose (140:109, or 1.3:1), this picture is dramatically reversed in imaginative prose (52:109, or 0.5:1). The findings for imaginative prose are not unexpected, in light of the distributions noted for LL. The fictional texts in LOB contain a substantial amount of dialogue and, given the comparatively high frequency of pseudo-clefts in speech, we would expect this to boost their overall numbers in fiction. Reversed pseudo-clefts are almost entirely restricted to passages of dialogue in fictional texts, with authors exploiting the fact that this construction is one that is favoured in informal speech. Thus the ratio of basic to reversed pseudo-clefts varies from 1.3:1 in Humour (R), where there is little dialogue, to 0.7:1 in General Fiction (K), where there is more dialogue, to 0.3:1 in Science Fiction (M) and Adventure and Western Fiction (N), where there is a high proportion of dialogue. A further possible explanation for the increase in popularity of pseudo-clefts from informative to imaginative prose follows from the claim made by Rader (1982) that fiction makes extensive reference to the internal temporal and physical situation constructed by the author (and thus, in terms of context-dependency, has more in common with spoken genres than informative written genres).

Within the informative genres of LOB, where clefts are almost twice as popular as pseudo-clefts, the frequency of clefts appears to vary along a dimension that might be characterized as 'factual/descriptive' versus 'opinionative/rhetorical'. Consider the ranking of Categories A-J, ordered according to the frequency of cleft constructions, in Table 7.5.

A strong determining factor in the frequency of cleft constructions in informative prose appears to be the extent to which the function of a text is on the one hand primarily descriptive (the aim being to present factual material objectively), or on the other hand primarily 'opinionative' (the aim being to mount arguments, engage in persuasion, mix opinion with fact, and so on). Of the nine informative categories, H and B perhaps represent these two generalized functions in their most extreme forms. The relatively infrequent occur-

Table 7.5 Frequency of clefts in the informative categories of LOB: fact versus opinion

Text category		Text frequency per 10,000 words
H	(Miscellaneous documents, reports, etc.)	1.8
A	(Press: reportage)	2.7
J	(Learned and scientific writings)	5.1
E	(Skills, trades, and hobbies)	5.1
F	(Popular lore)	6.3
G	(Belles lettres, biography, essays)	6.8
D	(Religion)	6.8
B	(Press: editorial)	7.0
C	(Press: reviews)	7.6

rence of clefts in categories H and A reflects their purely informative or expository orientation. In Categories G, D, B and C, by contrast, the higher frequency of clefts reflects their more subjective orientation: the writer is offering opinions and typically, attempting to persuade the reader. Part of the 'rhetorical' flavour of these categories derives from the writer's concern with textual aspects of the message (a concern which would make the special intonational/informational properties of the cleft construction, as I have described them above, an attractive resource). In addition to this, there will often be an historical element in the opinionative genres, which would be out of place in H and A where the temporal framework is either 'universal' or 'immediate'. Categories J, E and F are intermediate, with factual account mixing readily with opinion and persuasion.

It is instructive to compare the findings reported here for the registers of LOB with those recorded by Prince (1978). As noted in Section 5.3.5 above, Prince classifies clefts into two types, 'stressed-focus *it*-clefts' (corresponding to Type 1 in the present study), and 'informative-presupposition *it*-clefts' (corresponding to Types 2 and 3 in the present study). She characterizes the stylistic distribution of the latter as follows (1978:899-900):

they tend to occur in formal, often written, discourse . . . they are frequent in historical narrative, or wherever the speaker wishes to indicate that s/he does not wish to take personal responsibility for the truth or originality of the statement being made. In a sense, then, they are similar to hedges - e.g. *It seems that, I think that,*

sorta – since both have the effect of reducing the speaker's responsibility. Hedges do this by weakening the statement, by presenting it as an already known fact. Thus the *it*-cleft is particularly well-suited to persuasive discourse.

Prince attributes these distributional findings to the communicative nature of the relative clause: 'the *that*-clause contains the "message" – but marked as a known fact, not as the speaker's assessment' (1978:904). Prince's discussion of 'known' information leads her to a consideration of Whorf's description of nomic verbal affixes in Hopi:

In both English and Hopi, speakers can indicate how much responsibility they wish to take for the validity of their statements. In Hopi, they do it obligatorily by suffixation; in English they do it optionally in a variety of ways – hedges, 'future auxiliary markers', *it*-clefts etc.

(p.901)

[known information is] a choice on the part of the speaker of a particular validity level that s/he wishes to ascribe to the utterance. Markers of this choice (corresponding to Whorf's nomic) include the *it*-cleft and particular word choices (*the fact that*, *factive predicates* etc).

(p.903)

It may be noted that Prince wavers, in her discussion of the generic distribution and nomic characteristics of 'informative-presupposition *it*-clefts', between specific references to this type of cleft and references to clefts in general. In so doing, she implicitly acknowledges that the nomic function of clefts may have a source other than merely the communicative function of the relative clause in the 'informative-presupposition' type. A more likely basis for the 'responsibility-weakening' function of clefts is in fact the impersonal associations of the predicating *it* *be* structure (noted in the previous section). The structural similarities between clefts, and constructions of the type *it is said that* . . . *it is well-known that* . . . *it is arguable that* . . . and so on, are apparent. It may be true that the nomic function is more salient in marked clefts; even so, it is more plausible to suggest that its source here is in the confluence of newness, presupposition and syntactic dependency, rather than in a construction-specific form of shared knowledge (that is, Prince's 'known information').

Prince's claims would lead us to expect that historico-persuasive (in our terms 'opinionative') texts would yield a higher ratio of

marked:unmarked clefts than non-historico-persuasive (in our terms 'descriptive') texts. However, as Table 7.1 indicates, no such pattern emerges in LOB. Departures from the average marked:unmarked ratio of 2.5:1 for the nine informative categories are random (with H and A, for instance, having a higher ratio of marked clefts than B and C). The distributional findings for the informative LOB categories, as reported above, suggest that the generic tendencies noted by Prince for marked clefts in fact apply to clefts in general.

7.4 DISTRIBUTION OF CLEFTS AND PSEUDO-CLEFTS ACCORDING TO INFORMATIONAL PROPERTIES

In this section I shall consider the distribution of clefts and basic pseudo-clefts according to their informational classification. As was pointed out in Section 5.3.4 above, the classification of givenness that was designed for the analysis and classification of basic pseudo-clefts does not turn out to be relevant to reversed pseudo-clefts. In considering the communicative function of reversed pseudo-clefts it is important to observe the consistency with which they select their highlighted element/theme from a single, small, syntactic class. For this reason, discussion of the distribution of reversed pseudo-clefts, in terms of their communicative roles, is reserved until the following section.

(i) *Clefts*

In the combined corpus, as Table 5.3 indicates, 36.0% of all clefts are unmarked, and 64.0% marked (of the latter 34.6% are Type 2, and 29.4% are Type 3). The most striking difference between LL and LOB occurs with Type 3 clefts, which are considerably more popular in the written corpus. In this regard it might be noted that, while Prince does not distinguish between the two types of marked cleft, most of the examples that she provides of the 'informative-presupposition' class (which 'tend to occur in formal, often written, discourse', p.899) are in fact Type 3 clefts. The examination of genre categories below confirms that Type 3 clefts are preferred in formal, learned writing (Categories G and J in LOB) and in monologic speech. Another notable difference between LL and LOB occurs with unmarked clefts, which are relatively more popular in the spoken corpus. Perhaps one might have expected a high frequency of unmarked clefts in LOB, given the latent unmarkedness of intonation in writing. However, intonation seems to play a role of lesser importance than co-textual considerations in the determination of informativity with clefts (and probably in general) in written texts. It follows,

from the mere fact that clefts tend to have longer highlighted elements in writing than speech (see Section 7.6 below), that their highlighted elements are more likely to contain non-cotextually-recoverable information.

The spoken corpus evidences several clear trends (see Table 7.6). Unmarked clefts are most strongly represented in the private and dialogic subgroups (with a frequency of 1.9 per 10,000 words in Categories A-C and in Categories A-E). Their popularity drops in public texts (1.5 in Categories D-H), and even further in public monologues (0.9 in Categories F-H). This trend is reversed with Type 3 clefts, whose popularity increases from 0.8 in both private and dialogic texts, to 1.2 in public texts, and further to 1.4 in public monologues).

In LOB (see Table 7.6) the most striking contrast is in the relative popularity of unmarked clefts in the informative categories (1.5) and imaginative categories (3.2). On the other hand, both types of marked cleft are more popular in the informative than in the imaginative categories. As mentioned above there is, contrary to Prince's (1978) findings, no clear pattern of correlation in the informative categories between unmarked/marked cleft selection and genre categories (such as 'descriptive', 'opinionative', 'press').

The most significant register variables determining unmarked/ marked cleft selection seem to have been addressee-proximity, formality, and fictionality. The more directly interactive the text and the greater the informality, the higher the probability of unmarked cleft selection. Why should this be so? The explanation probably lies in the relative overall informativity of unmarked and marked clefts. The newsworthiness of an unmarked cleft is concentrated in its highlighted element (which, it may be noted, is shorter in the typical instance than the relative clause: see Section 7.6 below). By contrast the 'news' of the marked cleft is located, in the case of the Type 2 version in its relative clause, and in the case of the Type 3 version, in the highlighted element as well. It follows that, in terms of the lengths of their informative components, marked clefts are more oriented towards informativity than unmarked clefts (a claim supported by a comparison of the lengths of their informative components). This may explain, at least in part, why they are more favoured than unmarked clefts in formal, non-interactive, technical genres.

It is furthermore revealing to look more closely at the type of informativity that is characteristic of unmarked and marked clefts. As Table 7.6 shows, contrastive newness is more in evidence in unmarked than marked clefts. (In LL, 40.8% of the unmarked clefts

Table 7.6 Informational types of clefts in LL and LOB*

Cleft type	Subgroup	LL				LOB			
		A-C Private	D-H Public	Total	A-E Dialogue	F-H Monologue	Total	A-J Informative	K-R Imaginative
U	1a	19	7	26	25	1	26	39	32
N	1b	8	2	10	10		10	14	18
M	1c	10	9	19	16	3	19	35	11
A	1d	16	5	21	18	3	21	26	20
R	E								
K									
E	1	53	23	76	69	7	76	114	81
D	Total	(1.9)	(1.5)	(1.7)	(1.9)	(0.9)	(1.7)	(1.5)	(3.2)
T	2a	35	19	54	44	10	54	133	34
Y	2b	11	2	13	12	1	13	1	2
M	2c	1	3	4	1	3	4	16	2
P	2d		1	1		1	1		1
A	E								
R									
K	2	47	25	72	57	15	72	150	38
E	Total	(1.7)	(1.6)	(1.7)	(1.6)	(1.9)	(1.7)	(2.0)	(1.5)
T	3a	7	14	21	12	9	21	111	38
Y	3b	1	1	2	2		2	1	1
M	3c	9	3	12	10	2	12	22	5
P	3d	4		4	4		4	5	
E									
D	3	21	18	39	28	11	39	138	44
	Total	(0.8)	(1.2)	(0.9)	(0.8)	(1.4)	(0.9)	(1.8)	(1.7)

* () = tokens per 10,000 words

— that is, 1b + 1d — occurs with contrastive rather than freshly new highlighted items; in LOB the proportion is 40.0%. By contrast, in LL 28.8% of the marked clefts — that is, 2b + 2d + 3b + 3c + 3d — contain contrastive information; in LOB the proportion is only 9.7%.) Given that one might expect contrastiveness to be a feature more germane to interactive texts (rich as they are in mood and polarity variation, reflecting the existence of non-sharing contrasts in experience and point of view) than non-interactive texts (note the relative popularity of basic pseudo-clefts with 'opposite antecedence' in conversational and fictional texts: see below), the popularity of unmarked clefts in casual conversation is not unexpected. But why should unmarked clefts be so relatively popular in fiction? After all, fictional writing is no more interactive than non-fictional writing (if we can exclude from consideration the complicating exception of fictional dialogue: indeed the popularity of clefts in fiction *cannot* be explained, as can that of pseudo-clefts, in terms of their frequency in dialogue). Nevertheless the general level of formality of the fictional categories of LOB is intermediate between that of the relaxed conversational texts of LL and the expository texts of LOB. In many cases the style might be characterized as 'chatty', providing an appropriate context for the expression of contrasts. It therefore follows, from their distribution in LL, that we might expect unmarked clefts to be more popular in fiction than marked clefts. The following typical examples from fictional texts give some idea of unmarked clefts at home in their semi-formal surroundings:

- (1) 'Perhaps the fishes'll grow as strong and virile as the Shoshone Indians. Still, it must have been a wonderful show.'
'He made a wonderful fool of himself,' Agnes said.
It was fate, in fact, that was making fools of all of us. I said before that the stage had been set — it awaited the last theatrical prop. (LOB K09.93)
- (2) Whispering tongues blamed her

When Rob Archer's wife Anne, disappears, the police believe it is because she is guilty of poisoning Bob's mother. Actually, however, she is being held prisoner by Vera Corbett. Mrs Archer's ex-maid. *It was Vera who killed Bob's mother and she knows Anne can give her away.* Arthur Hedley, Vera's boyfriend, also knows of her guilt, but he is too deeply involved to back out. (LOB L22.4.6)

Consider finally the register distribution of the two marked cleft

types. There is little that can be said about the generic distribution of Type 2 clefts. They are consistently well represented across the spoken genres, and in informative prose, but this popularity wanes in fiction. The distribution of Type 3 clefts, on the other hand, is largely complementary to that of Type 1 clefts. In other words, Type 3 clefts are preferred in formal, non-interactive texts. Why should this be so? The answer presumably lies largely in the high level of informativity of the Type 3 construction. It is this informativity which explains the potential of the construction for occurrence in text/stage-initial position (that is, in positions where few background assumptions can be made): see further Section 6.3.2 above. In this regard it is instructive to observe the contrast between Type 3 clefts and reversed pseudo-clefts. In the latter construction, low informativity correlates with a tendency to occur in text/stage-final position (see further Section 6.2.2 above).

(ii) *Basic pseudo-clefts*

The classification of givenness-types for basic pseudo-clefts that is developed in Chapter 4 above and exemplified in Chapter 6 is organized along a number of (sometimes intersecting) parameters. In this section the distribution of pseudo-clefts across the categories defined by these parameters is examined.

While in the spoken corpus (see Table 7.7) the number of basic pseudo-clefts with co-textual and contextual antecedents is almost identical (138:134), in the written corpus (see also Table 7.7) there are almost twice as many with co-textual antecedents (124:68, or 1.8:1). While this difference would seem to follow from claims regarding the relative context-dependency of speech and writing in Section 2.3 above, curiously, the figures for genres within the two corpora mostly fail to conform to the expected trends. Thus, contextual antecedence is not, as one might have anticipated, more common in the 'more speech-like' private conversational texts or in the total set of dialogic texts. Contextual antecedence was predictably less common in the informative genres of LOB, but its extremely poor showing in the imaginative genres (5 out of 52 tokens) is contrary to expectation.

(a) *Co-textual antecedence*

Within the category of co-textual antecedence, the four subcategories manifested by the intersection of the directness/indirectness and similarity/oppositeness parameters yield distributional differences both across and within the two corpora.

Table 7.7 Givenness and basic pseudo-clefts in subgroups of LL and LOB*

Corpus	Subgroups	Co-text			Context		
		Directly similar	Directly opposite	Indirectly similar	Indirectly opposite	Field	Tenor
LL	A-C Private	1.2 (33)	0.6 (16)	1.0 (27)	0.6 (16)	1.1 (32)	0.8 (23)
	D-H Public	1.3 (20)	0.6 (10)	0.9 (14)	0.1 (2)	1.9 (30)	2.5 (69)
	A-E Dialogue	1.3 (46)	0.7 (26)	1.0 (34)	0.5 (18)	1.5 (52)	1.1 (40)
	F-H Monologue	0.9 (7)	0.0 (0%)	0.9 (7)	0.0 (0%)	1.3 (10)	0.8 (6)
	Total	1.2 (53)	0.6 (26)	0.9 (41)	0.4 (18)	1.4 (52)	1.1 (46)
LOB	A-J Informative	0.1 (8)	0.2 (13)	0.6 (44)	0.2 (12)	0.3 (20)	0.4 (30)
	K-R Imaginative	0.1 (2)	0.6 (14)	0.6 (15)	0.6 (16)	0.1 (2)	0.1 (3)
	Total	0.1 (10)	0.3 (27)	0.6 (59)	0.3 (28)	0.2 (22)	0.3 (33)
				1.2 (124)			0.7 (68)

* Frequencies = tokens per 10,000 words; () = raw frequencies

To take the direct/indirect opposition first, it emerges (see Table 7.8) that while speech favours direct antecedence (by a ratio of 79:59, or 1.3:1), writing favours indirect over direct antecedence (by a ratio of 87:37, or 2.4:1). As mentioned in Section 6.1.2 above, these findings are broadly compatible with the findings of Prince in her (1981) study. In her comparison of two texts, one spoken and one written, analysed in terms of 'assumed familiarity', she found that the written text relied more heavily than the spoken on inferrability. Further study of such differences is required before we can know the answers to questions of the following type: Is it generally true that writers impose greater cooperative demands upon their readers, than speakers do upon their listeners? What is the relationship between the directness/indirectness distinction and register variables? Do direct/indirect ratios vary as writers proceed from first draft to final draft?

Table 7.8 Co-textual givenness and basic pseudo-clefts in subgroups of LL and LOB*

Corpus	Subgroups	Similar	Opposite	Direct	Indirect
LL	A-C Private	2.1 (60)	1.1 (32)	1.8 (49)	1.5 (43)
	D-H Public	2.2 (34)	0.8 (12)	1.9 (30)	1.0 (16)
	Total	2.2 (94)	1.0 (44)	1.8 (79)	1.4 (59)
LOB	A-J Informative	0.7 (52)	0.3 (25)	0.3 (21)	0.7 (56)
	K-R Imaginative	0.7 (17)	1.2 (30)	0.6 (16)	1.2 (31)
	Total	0.7 (69)	0.6 (55)	0.4 (37)	0.9 (87)

* Frequencies = tokens per 10,000 words; () = raw frequencies

Once again, there are some surprises in store when we examine subgroupings within the corpora. The informative categories of LOB demand, as one might expect, more work of their readers than the imaginative categories (the indirect:direct ratio in informative texts is 56:21, or 2.7:1; in imaginative texts 31:16, or 1.9:1). However in LL it is the private texts which evidence more indirectness (with a direct:indirect ratio of 49:43, or 1.1:1) than the public texts (with a direct:indirect ratio of 30:16, or 1.9:1).⁴

Turning to the similar/opposite opposition, we see (Table 7.8) that both speech and writing favour similarity, though the ratios are dissimilar (for speech the similar:opposite ratio is 94:44, or 2.1:1; for

writing it is 69:55, or 1.3:1). In light of these ratios the figures for subgroupings of the corpora hold some surprises. It is the public texts of LL which yield the stronger preference for similarity (with a similar:opposite ratio of 34:12, or 2.8:1, compared with a ratio of 60:32, or 1.9:1, for private texts). In LOB it is the informative texts in which similarity is best represented (with a similar:opposite ratio of 52:25, or 2.1:1, compared with a ratio of 17:30, or 0.6:1, for imaginative texts).

(b) Contextual antecedence

Within the category of contextual antecedence the order of preference for the three categories is, for LL: field (62), tenor (46), mode (26); for LOB it is tenor (33), field (22), mode (13) (Table 7.7).

The popularity of the field category in speech is predictable on several counts. Firstly, the presence of a pro-verb, which is a hallmark of the category, introduces an element of redundancy that will tend to be more at home in speech than writing. Secondly, the fact that the pro-verb is, in Halliday's terms, 'material' rather than 'relational', 'mental', or 'verbal', suggests that it will be more often represented in speech (in view of the field differences between spoken and written genres discussed in Section 2.3 above).

The relative popularity of the tenor category in writing is a little more difficult to explain. Perhaps the reason is that there is less need for people to articulate their emotions and reactions in speech, where these may alternatively be conveyed paralinguistically.

In both corpora the mode category is the least frequent, accounting in both bases for only about one-fifth of tokens in the contextual categories.

7.5 HIGHLIGHTED ELEMENTS AND REGISTER

7.5.1 *Explaining the correspondences*

In Chapter 4 above it was noted that there are striking differences between cleft and pseudo-cleft constructions in the syntactic classes of their highlighted elements. Pseudo-clefts favour nominal elements (in the basic construction: noun phrases, finite clauses, non-finite clauses; in the reversed construction: predominantly demonstratives). Clefts are by contrast extremely flexible, accepting as highlighted element not only any item that has a representational function

in the clause (and which can be thematic in the corresponding non-cleft), but also 'zero'.

It is tempting to consider the implications of these findings for the broad distributional differences between the constructions. In one sense, the existence of a relationship has already been established implicitly: it has been shown that the choice of highlighted element is related to communicative differences (such as the textual/ideational highlighting distinction), and that textual structure (that is, questions of thematicity and informativity) is related to register distribution. By inference, then, one might posit the existence of links between choice of highlighted element and register distribution.

However it is possible to mount a more explicit case for the existence of correspondences between formal properties of clefts and pseudo-clefts, and their generic distribution. Several lines of argument are pursued below.

(i) *Thematic flexibility*

Clefts and pseudo-clefts exhibit 'thematic flexibility', but their flexibility is of different kinds and corresponds to structural differences between the constructions. The thematic flexibility of pseudo-clefts has a 'syntagmatic' orientation: the equative structure with thematic nominalization enables all possible distributions of clause elements into the theme and rheme (and, furthermore, inversion of the constituents realizing these functions). The thematic flexibility of clefts has a 'paradigmatic' orientation: the internal predication selects a single element for textual prominence. Here the thematic inflexibility at the syntagmatic level is counterbalanced by flexibility of a paradigmatic kind (in the wide range of syntactic classes that may be theme-predicated). The next step in the argument concerns the congruity between these different types of thematic flexibility, and the different patterns of complexity typical of spoken and written language. I have already noted (Section 2.3) that speech is typified by complexity of a dynamic, or 'choricographic' nature, writing by complexity of a synoptic, or 'crystalline' nature. On the basis of such differences one might predict that the syntagmatic flexibility of pseudo-clefts would be more germane to speech (where in fact pseudo-clefts are more frequent than clefts) and the paradigmatic flexibility of clefts more germane to writing (where in fact clefts are more frequent than pseudo-clefts).

(ii) *Language processing*

Corresponding to the difference in thematic flexibility between basic pseudo-clefts and clefts are certain processing differences that may be deduced from tokens in LL. In basic pseudo-clefts, the progression

from explicitly represented 'non-news', to 'news', gives the construction a linearly developmental character. The concern of the thematic relative clause with presenting material accessible from co-text or context offers the speaker an extended opportunity to formulate the 'message'. Where the relative clause contains contextually recoverable material (as it does considerably more often in speech than writing: see previous section), the content of the highlighted element is typically unpredictable from that of the relative clause. Thus basic pseudo-clefts with field-recoverability (for example *What happened was . . .*; *what he did was . . .*), tenor-recoverability (for example *What surprised me was . . .*) and mode-recoverability (for example *What this means is . . .*) are particularly prone to being extended considerably beyond the copula. By contrast in clefts the speaker is obliged (except in the case of Type 2 clefts) to present news at the onset, suggesting that clefts will tend to be encoded with much of their content 'pre-formulated'. It is presumably as a consequence of these processing differences that system-deviance (as discussed and exemplified in Section 3.4 above) is common with basic pseudo-clefts in LL, but rare with clefts.

It is also possible to offer a language-processing explanation for the fact that reversed pseudo-clefts are more popular than basic pseudo-clefts in speech (particularly unpremeditated varieties). The rhematic content of the reversed construction is even less predictable from the content of its theme than is the case with the basic construction. Typically, as we have seen, the theme of reversed pseudo-clefts is a referentially vague-weakly informative demonstrative. The experiential role of this demonstrative does not need to be determined by the speaker until the commencement of the rheme (as in *This is the one . . .*, *This is why . . .*, *This is the place . . .*, *This is how . . .*, and so on).

(iii) *Indirectness and deference*

A further possible explanation for the register distribution of basic pseudo-clefts relates to their formal semantics. Consider the following example:

- (3) B [ə:m] – I did most of the documentation for volume THREE #
 a Professor Leegate
 B YES #
 a is this the Canadian girl
 B ((CAROLINE #, CAROLINE #))
 a yes yes (clears throat)

- B I . did most of the documentation research for volume THREE # – FOR her # – [ə] it . helped me out of a bad SPOT #
 a (laughs)
 B [ə:m, ʔ w] *what I wanted to ask you was THIS* # I'll be [ʔə] PÈRFEKTLÏ FRÁNK # . [ə:m] – the bibliography has GÖNE # about as FÁR # as I can take it on my ÖWN # – that is to SÁY # – [ə:] – in order to . COMPLÈTE it # – I will have to visit . the MÁJOR # resources – in the UNITED STÁTES # . and [ə] several in ÈUROPE # – and [ə] – YÖU KNÖW # my own personal finances are ((not 3 sylls))
 a well snre (LL S.2.1.45)

Speaker B's tone here might best be characterized as deferential and courteous. The request that he formulates, with hesitation and circumlocution, in his last turn in this extract, is initiated with a basic pseudo-cleft. This choice of construction is most apt, enabling him to avoid a direct assertion of his desire to solicit a favour from Speaker A (compare the increasing degrees of remoteness in the series: 'I want to ask you this' > 'I wanted to ask you this' > 'What I wanted to ask you was this'). The basic pseudo-cleft construction enables Speaker B politely to relegate his wish to the status of presupposed background knowledge, while at the same time highlighting it as the theme.

It is this indirectness (realized in the form of a thematically prominent presupposition which the addressee is expected to cooperatively assume) that, as a feature of the 'interpersonal semantics' of the basic pseudo-cleft construction, provides an explanation for its popularity in spoken genres.

7.5.2 *Distribution of clefts and pseudo-clefts according to highlighted elements*

There are variations in the register distribution of cleft and pseudo-cleft constructions with particular highlighted elements, which reinforce the hypothesis that relationships exist between syntax and register with these constructions.

(i) *Clefts*

Numbers for clefts with highlighted noun phrases are comparable in the two corpora, occurring at a rate of 2.5 tokens per 10,000 words in LL (and accounting for just over half (57.2%) of all cases) and at a

rate of 2.7 tokens per 10,000 words in LOB (and accounting for just under half of all cases (48.1%); see Table 7.9). Nor are there any radical departures from this pattern in subgroupings of the corpora.

Clefts with highlighted prepositional phrases, however, evidence striking variation, being over four times (4.3:1) more popular in writing than speech. The generic distribution of clefts with highlighted prepositional phrases is explicable in terms of the thematic demands of writing. It seems plausible to suggest that writing, by contrast with speech, depends for its texture more on thematic than informational patternings (and particularly opinionative writing, where many writers make a conscious attempt to exercise verbal artistry). A construction which readily permits the thematization of prepositional phrases provides a valuable resource for writers seeking flexibility of theme-selection.

Clefts with zero-highlighted item are more popular in speech (1.0 tokens per 10,000 words) than in writing (0.7), and within the spoken genres they are more common in private texts (1.2) than in public texts (0.7), and more common in dialogic texts (1.1) than monologic texts (0.8).

The numbers for clefts containing the remaining four types of highlighted element (as presented in Table 7.9) are arguably too small (representing only 14.3% of the combined corpus) for us to be able to talk of systematic differences in register distribution. Of the four, the most radical difference between the two corpora occurs with highlighted adverb phrases, which are considerably more common in writing than speech. The order of the difference is similar to that found with prepositional phrases (not unexpectedly, given the semantic similarity between prepositional phrases and adverb phrases).

(ii) *Basic pseudo-clefts*

It is instructive to consider the figures for highlighted elements with basic pseudo-clefts (see Table 7.10) in light of the discussion of lexical density and grammatical intricacy in Section 2.3 above. That basic pseudo-clefts should select, as highlighted element, noun phrases more often in writing than speech, and finite clauses more often in speech than writing, is compatible with Halliday's contention (and supported by Biber's findings) that speech is less lexically dense and more grammatically intricate (characterized by more subordination) than writing. Within subgroupings of LL, predictably, basic pseudo-clefts with highlighted noun phrases are more frequent in public varieties (2.5 per 10,000 words) than private varieties (1.3), while those with highlighted finite clauses occur with comparable frequency in the two varieties (3.2 public; 3.0 private). Basic pseudo-clefts with

Table 7.9 Highlighted elements of clefts in subgroupings of LL and LOB*

Corpus	Subgroups	NP	PP	Ø	FIN CL	ADV P	NF CL	ADV P	Total
LL	A-C (Private)	2.4 (68)	0.3 (9)	1.2 (33)	0.4 (10)	0.0 (1)			4.3 (121)
	D-H (Public)	2.5 (39)	0.4 (6)	1.7 (11)	0.5 (7)	0.1 (2)	0.1 (1)		4.3 (66)
	A-E (Dialogue)	2.4 (86)	0.3 (11)	1.1 (38)	0.4 (15)	0.1 (3)	0.0 (1)		4.3 (154)
	I-H (Monologue)	2.6 (21)	0.5 (4)	0.8 (6)	0.3 (2)				4.1 (33)
	Total	2.5 (107)	0.3 (15)	1.0 (44)	0.4 (17)	0.1 (3)	0.0 (1)		4.3 (187)
LOB	A-J (Informative)	2.4 (183)	1.6 (117)	0.7 (50)	0.3 (22)	0.3 (25)	0.1 (5)		5.4 (402)
	K-R (Imaginative)	3.5 (89)	1.2 (30)	0.6 (15)	0.5 (12)	0.6 (16)		0.0 (1)	6.5 (163)
	Total	2.7 (272)	1.5 (147)	0.7 (65)	0.3 (34)	0.4 (41)	0.1 (5)	0.0 (1)	5.7 (565)

* Frequencies = tokens per 10,000 words; () = raw frequencies.

highlighted noun phrases are equally frequent in the informative and imaginative varieties of LOB (0.8). A similar pattern occurs with pseudo-clefts with highlighted finite clauses (informative varieties 0.8; imaginative 0.7).

Table 7.10 Highlighted elements of basic pseudo-clefts in subgroups of LL and LOB*

Corpus	Subgroups	NP	FIN CL	NF CL	PP	Total
LL	A-C (Private)	1.3 (36)	3.0 (84)	1.4 (40)	0.0 (1)	5.8 (161)
	D-H (Public)	2.5 (39)	3.2 (49)	1.4 (21)	0.1 (2)	7.2 (111)
	A-E (Dialogue)	1.8 (65)	3.3 (118)	1.6 (57)	0.1 (2)	6.8 (242)
	F-H (Monologue)	1.3 (10)	1.9 (15)	0.5 (4)	0.1 (1)	3.8 (30)
	Total	1.7 (75)	3.1 (133)	1.4 (61)	0.1 (3)	6.3 (272)
LOB	A-J (Informative)	0.8 (60)	0.8 (58)	0.3 (21)	0.0 (1)	1.9 (140)
	K-R (Imaginative)	0.8 (19)	0.7 (17)	0.6 (15)	0.0 (1)	2.1 (52)
	Total	0.8 (79)	0.8 (75)	0.4 (36)	0.0 (2)	1.9 (192)

* Frequencies = tokens per 10,000 words; () = raw frequencies

Table 7.11 Highlighted elements of reversed pseudo-clefts in subgroups of LL and LOB*

Corpus	Subgroups	that	this	Total
LL	A-C (Private)	6.4 (179)	2.0 (55)	8.4 (234)
	D-H (Public)	2.9 (45)	1.1 (17)	4.0 (62)
	A-E (Dialogue)	5.9 (210)	1.9 (67)	7.8 (277)
	F-H (Monologue)	1.8 (14)	0.6 (5)	2.4 (19)
	Total	5.1 (224)	1.7 (72)	6.8 (296)
LOB	A-J (Informative)	0.6 (43)	0.4 (32)	1.0 (75)
	K-R (Imaginative)	3.3 (84)	0.4 (9)	3.7 (93)
	Total	1.3 (127)	0.4 (41)	1.7 (168)

* Frequencies = tokens per 10,000 words; () = raw frequencies

(iii) *Reversed pseudo-clefts*

The relative frequency of the two main highlighted elements that occur with reversed pseudo-clefts, *that* and *this*, remains constant across the two corpora, *that* being three times more popular than *this*.

This ratio is maintained with remarkable consistency, through the subgroups (private/public, dialogue/monologue) of LL (see Table 7.11). In LOB the popularity of *that* falls in informative genres and rises in imaginative genres. The reason for the latter result, bearing in mind that most reversed pseudo-clefts in fiction occur in dialogue, is probably that writers exaggerate the speech-tendency for *that* to outnumber *this* in reversed pseudo-clefts. The result for the informative genres is possibly to be explained in terms of the shorter 'reach' of anaphora in writing, making *this* a suitable choice of highlighted demonstrative in many cases.

7.6 CLEFTS, PSEUDO-CLEFTS, AND WORD LENGTH

In this section I shall examine the possibility, suggested by Prince (1978), that there exists a close relationship between the word length of pseudo-cleft and cleft constructions, and their informativity. It is readily admitted that calculating the size of the grammatical units in question in terms of running words is open to criticism. It would undoubtedly have been more revealing to use the phrase as a basis for such calculations. Unfortunately, the analysis that such an approach demands is beyond the scope of the present study.

As Table 7.12 indicates, Prince's figures and those derived from the present study are in broad agreement. The word length of units was calculated in the same way as in Prince's study, in order to facilitate comparison. Thus, the figures under 'highlighted element' ('focus', in Prince's terms) refer to the number of words in the constituent following the copula. The figures under 'relative clause' ('presupposition', in Prince's terms) refer, in the case of pseudo-clefts, to the number of words following *what*, *all*, *the thing that*, and so on, and in the case of clefts, to the number of words following *that*, *which*, and so on. In other words, the total word count for each construction should be equivalent to that of the corresponding non-cleft. Reversed pseudo-clefts are not included in the Prince study.

Prince observes that her pseudo-cleft figures

seem intuitively correct if we follow the usual custom of equating logical presupposition with old/given/known information. That is, it seems natural to use as few words as possible to refer to something known, and to present new information in greater detail; this is in line with one of Grice's Maxims of Manner: 'Be brief (avoid unnecessary prolixity)' (1975:46)

(p.886).

Table 7.12 Clefts, pseudo-clefts, and word length: a comparison of Prince and Collins

Construction	Researcher	Speech		Writing	
		Highlighted element	Relative clause	Highlighted element	Relative clause
(Basic) pseudo-clefts	Prince	11.4	4.6	19.3	5.3
	Collins	13.5	4.6	15.8	5.0
Clefts	Prince	2.5	4.2	6.1	10.5
	Collins	4.0	7.2	5.1	9.9

The suggestion that the length of the highlighted element and relative clause reflects their informativity merits closer scrutiny. Tables 7.13 and 7.14 present word length averages for the major informational types of basic pseudo-clefts and clefts.

Table 7.13 Basic pseudo-clefts: informativity and word length in LL and LOB

Corpus	Context			
	Directly similar	Directly opposite	Indirectly similar	Indirectly opposite
LL	2.8 : 11.9	5.3 : 16.2	5.5 : 12.4	5.8 : 11.9
LOB	6.3 : 22.1	4.4 : 14.6	6.4 : 13.4	4.9 : 13.8
Total	3.4 : 13.6	4.8 : 15.4	6.0 : 13.0	5.3 : 13.1
	5.0 : 13.6			

Corpus	Context		
	Field	Tenor	Mode
LL	4.2 : 13.9	5.7 : 17.5	3.5 : 13.6
LOB	5.2 : 15.0	3.8 : 20.7	4.8 : 21.8
Total	4.4 : 14.2	4.9 : 18.9	3.9 : 16.2
	4.5 : 16.4		

The figures in these tables clearly support Prince's claims. In all the givenness-types of basic pseudo-clefts the relative clause outstrips the highlighted element in length (with ratios varying between 1:2.2 for the 'indirectly similar' category to 1:4.8 for the 'mode' category).

Table 7.14 Clefts: informativity and word length in LL and LOB

Corpus	Type 1	Type 2	Type 3
LL	5.2 : 3.0	2.0 : 9.8	5.0 : 9.4
LOB	5.6 : 3.3	2.2 : 13.3	7.3 : 13.6
Total	5.5 : 3.2	2.2 : 12.3	6.9 : 12.8

Within the co-textual categories indirect relative clauses are longer than direct relative clauses (6.0 and 5.3 words for 'indirectly similar' and 'indirectly opposite' respectively, as opposed to 3.4 and 4.8 words for 'directly similar' and 'directly opposite' respectively). This finding is as one might have expected, inferrability being less 'non-new' than directness.

With unmarked clefts, in which the highlighted element is new and the relative clause given, the former is predictably longer than the latter. By contrast, the two subtypes of marked clefts, which I have argued (in Section 5.3.5) are alike in the informativity of the relative clause (a contention supported here by the comparability in length of their relative clauses: 12.3 words for Type 2, 12.8 words for Type 3), have highlighted elements of interestingly different lengths. In the case of Type 2 clefts, this constituent is only 2.2 words on average, a finding supportive of the contention that it contains information which is given (because non-contrastively anaphoric or deictic). In the case of Type 3 clefts, this constituent is 6.9 words on average, corresponding to the fact that it contains new information (or perhaps more appropriately 'semi-new', typically a scene-setting adjunct).

It is furthermore pertinent, in view of the apparent relationship between informativity and word length, to consider whether there are any correspondences between word length and the register distribution of cleft and pseudo-cleft constructions. Starting with the broad distinction between speech and writing, it can be seen from Table 7.15 that clefts and pseudo-clefts are consistently longer in writing. (A comparison of the totals for LL and LOB reveals that, with all three constructions, both the highlighted element and relative clause evidence a proportionally similar increase as we move from speech to writing). These findings lend support to Halliday's claims regarding the differences between speech and writing (see Section 2.3). The relative grammatical simplicity and lexical density of writing ensure that clauses will be longer on average in that mode than in speech, with its grammatical complexity and lexical sparseness.⁵

Within subgroupings of the two corpora, increased clause length

Table 7.15 Word length of highlighted element/relative clause in clefts and pseudo-clefts in LL and LOB

Corpus	Subgroup	Clefts		Pseudo-clefts			
		Highlighted element	Relative clause	Basic		Reversed	
				Highlighted element	Relative clause	Highlighted element	Relative clause
LL	A-C (Private)	4.0	6.4	12.4	4.1	1.1	4.7
	D-H (Public)	4.2	10.0	15.0	5.4	1.3	6.9
	Total	4.0	7.2	13.5	4.6	1.2	5.2
LOB	A-J (Informative)	5.5	11.2	18.0	5.2	2.5	9.2
	K-R (Imaginative)	4.0	6.8	10.1	4.8	1.6	7.0
	Total	5.1	9.9	15.8	5.1	2.0	8.1

tends to correlate with an increase in abstractness, formality and context-independence. Accordingly, both clefts and pseudo-clefts are longer in the 'public' subgroup of LL than in the 'private' subgroup, and longer in the 'informative' subgroup of LOB than in the 'imaginative' subgroup (see Table 7.15). Furthermore, the greatest increases tend to occur in the more informative constituent. In the case of clefts this constituent, the relative clause (in the more frequent marked construction), increases more than the highlighted element as we move from private to public texts in LL, and from imaginative to informative texts in LOB. The same is true of reversed pseudo-clefts, where again it is the relative clause that is typically more informative. In basic pseudo-clefts, where the highlighted element is more informative than the relative clause, this element shows the greater increase from the imaginative to the informative genres of LOB. The only exception to the pattern is the private and public genres of LL, in which the relative clauses of basic pseudo-clefts show a marginally larger increase than the highlighted elements.

8 Conclusion

The adoption of a corpus-based approach to the study of cleft and pseudo-cleft constructions has proved to have a number of advantages. Such an approach has enabled distributional tendencies to be investigated empirically across a broad range of genres. The figures thus derived are, it has been demonstrated, reliable to grammatical and informational/thematic features. Furthermore, study of the functions and distribution of clefts and pseudo-clefts across a variety of spoken and written registers provides a potential source of hypotheses about the textual organization of discourse.

Another advantage is that the 'total-accountability' principle that is fundamental to the corpus-based approach provides a safeguard against the idiosyncratic biases, accidental omissions and distortions which are an ever-present danger in an introspective approach. (For example, one characteristic of clefts that the present study has confirmed is thematic flexibility in the choice of highlighted element.) And this property has been shown to relate integrally to structural attributes of the cleft construction. In view of these findings one might expect that the conclusions of those scholars who concentrate, in discussions of clefts, upon those with highlighted noun phrases alone (for example Hornby 1971, Harries-Delisle 1978, Atlas and Levinson 1981) will be limited and possibly distorted. As another example, consider the effects of embracing the total-accountability principle in the description of pseudo-clefts. Once we have acknowledged the existence of reversed pseudo-clefts, it is difficult subsequently to exclude them from attention, in the way that for example Prince (1978:905) does, noting: 'One different but closely related construction that should be studied is the inverted WH-cleft', and as Higgins (1979:6-7) does, saying: 'I shall not in this essay provide a treatment of "inverted" pseudo-cleft sentences . . . [They] show rather different properties from pseudo-cleft sentences, being

more akin to cleft sentences in some ways, and they require a separate investigation.'

Using LL and LOB as a database, it has been possible to provide quantitative support for descriptive findings at all levels of analysis. Correlations between grammatical structure, informational/thematic function and register distribution have been noted and are claimed to provide partial evidence in favour of the principle – axiomatic to functional theories of language – that there is a systematic relationship between language structure and language function. There follows a brief summary of the main claims and findings of the study.

In defining the class of pseudo-clefts and clefts it is necessary, we have seen, to distinguish them from constructions which have a similar sequence of elements but which are to be interpreted differently. While genuine pseudo-clefts can be readily distinguished from superficially similar attributive constructions using the criterion of reversibility, delimiting the class of clefts is a less straightforward matter. Proverbial sentences of the type 'It's a poor heart that never rejoices', and constructions with referential *it*, are certainly to be rejected, but there are several reasons for accepting the borderline constructions which Declerck labels 'predicational clefts'. Despite the fact that their non-cleft agnates are attributive, they are clearly identifying in structure.

The class of pseudo-clefts, it has been argued, comprises three subclasses. In addition to the canonical construction with nominal relative clause introduced by *what*, I have accepted those with relative clause introduced by the other *wh*-items of English, and those with relative clause introduced by *the* in conjunction with their pro-form equivalents. In the latter case the criterion that has been used to determine the possible extent of modification of the pro-form head is that of unclausability (that is, the availability of a non-cleft agnate). A further subclass of pseudo-clefts (included on the grounds of proportionate semantic relationships) is that comprising identifying constructions with an *all*-headed relative clause.

Linguists have tended to concentrate their attention on selected syntactic and semantic characteristics of clefts and pseudo-clefts, as determined by their specific objectives. Thus, those whose primary concern has been to define formally the relationships between these constructions and their non-cleft counterparts have concentrated upon so-called 'connectedness' phenomena. Those whose interest is in formal semantics have focused on the implicational and presuppositional features that differentiate cleft and pseudo-cleft constructions from their non-cleft counterparts. Those whose interest is in the

informational/thematic properties and discourse functions of clefts and pseudo-clefts have considered the range of syntactic classes of their highlighted elements, and the functions that they occupy in the non-cleft version (these functions being equivalent to those of the relative item). Because it is the latter concern that constitutes the major focus of the present study, considerable space has been devoted to a description of the selection of highlighted elements.

(Pseudo-clefts and clefts have been found to differ significantly in the class and function of their highlighted elements. The present study disconfirms Halvorsen's (1978:88) claim that 'The restrictions on pseudo-cleft foci are in general much less strict than the restrictions on cleft foci', and supports Prince's (1978:885) claim that 'the only significant overlap concerns focused NP's'. As noted above, a noun phrase is selected as highlighted element in a third of pseudo-clefts, in one half of clefts, and in the vast majority of reversed pseudo-clefts. Pseudo-clefts strongly favour nominal elements (in the case of the basic construction, both phrasal and clausal: only one in twenty of the highlighted finite clauses are adverbial, rather than content, clauses). The preference is very strong with reversed pseudo-clefts, where *this* and *that* are selected even when the highlighted element functions as adjunct, in preference to *here*, *there*, *then*, *in this way*, and so forth. The findings for pseudo-clefts are compatible with the claim that pseudo-clefts encode an experiential form of thematic highlighting, deriving from the (reversible) equation of two elements in an identifying relationship.

(By contrast, clefts highlight virtually any item which is able to be thematized in the corresponding non-cleft. (Not only noun phrases, but also prepositional phrases, finite clauses, adverbial phrases, and zero (where the cleft thematizes an indication of tense, modality, polarity, and so on, with all items having a representational function appearing in the relative clause), are represented in significant numbers. The fact that there are, amongst these items, several (zero and certain prepositional phrases) which could only be thematic in a theme-predicated structure and not in a pseudo-cleft (because they cannot be substituted by a *wh*-item) reinforces the claim that clefts display a different form (textual) of thematic prominence.)

(Of the syntactic functions occupied by the highlighted elements of clefts and pseudo-clefts, the three that are primary – subject, object, and adjunct – have strikingly different weightings in basic pseudo-clefts, reversed pseudo-clefts, and clefts.) The functions that are most often thematic in ordinary non-cleft declaratives (subject and adjunct) are favoured in clefts, suggesting that their typical function is

to imbue an already thematic item with further prominence (through predication). With basic pseudo-clefts the popularity of highlighted objects is predictable from English word order, but the popularity of subjects (*vis-à-vis* adjuncts) reflects a preference for participant-related functions. With reversed pseudo-clefts the preference for highlighted objects and adjuncts over subjects is the reverse of the situation with clefts, suggesting that reversed pseudo-clefts achieve thematic prominence by selecting functions that are most unlikely to be thematic in ordinary declaratives.

(The key to the analysis of cleft and pseudo-cleft constructions, it has been argued, is informational/thematic. Their primary function, as the name 'cleft' suggests, is to enable subsets of elements to be grouped into two parts in an almost unlimited number of ways. One aspect of this thematic flexibility is the possibility of incorporating all elements with ideational function into the second part (that is, the rheme), enabling thematization of interpersonal items. Thus in a pseudo-cleft beginning *What might have happened* . . . , the pro-verb directs the addressee towards the rheme for a specification of the process, while the modality and tense items associated with it are presented as thematic. Similarly, in a cleft beginning *It may have been that* . . . , none of the elements in the theme has an ideational function. I have argued that there is a difference in the way that pseudo-clefts and clefts generate thematic prominence. In the case of pseudo-clefts the prominence is ideational in orientation, deriving from the structural equation of the highlighted element and relative clause. In clefts the prominence is textual in orientation, deriving from the structural device of predication, which introduces the theme as complement to the non-referential subject *it*. Clefts are not reversible, as are pseudo-clefts, so that the emphasis falls less upon the identity between the two parts than upon the predication of one part in the structure (the theme).

There are informational, as well as thematic, differences between pseudo-clefts and clefts. Basic pseudo-clefts display a consistently close mapping of theme (along with presupposition), and givenness on to the subject relative clause, which is presented to the addressee as representing information that s/he should be prepared to accept as non-controversially recoverable. Basic pseudo-clefts thus offer the speaker a means of specifying precisely, before the announcement of the 'message', the background knowledge to which the addressee is expected to have access. The source for this knowledge may be co-textual or contextual. As we have seen, whereas co-textual and contextual sources are almost identical in number in LL, in LOB

there were almost twice as many co-textual. These findings presumably reflect general mode differences between speech and writing.

Co-textual recoverability may be quite straightforward, or it may be less direct, requiring the addressee to infer cooperatively an 'antecedent'. Direct recoverability is favoured in LL, inferable recoverability in LOB. It would be interesting to explore whether these differences relate to a general trend for writers to impose greater cooperative demands on their addressees than speakers.

* (Contextual recoverability may be of three types, deriving from the three variables that determine register. 'Field-antecedents' are retrievable from the setting of relevant events and happenings within which the language is functioning. 'Tenor-antecedents' are those that focus on the speaker/writer's emotions, reactions, thoughts and so on, which are presented as a legitimate concern of the cooperative addressee. 'Mode-antecedents' are those which focus on the mode of communication itself, and are thus intrinsically metalinguistic.)

(Whereas in basic pseudo-clefts it is the highlighted element that conveys the 'news', this element in the reversed construction is typically represented by a demonstrative with extended-text referential function (and thus inherently given, unless contrastive). (The typically text referential function of the highlighted element, and the exclusive equation of this element with the backgrounded material in the relative clause, give rise to a special 'internal-referencing' discourse function. With their capacity to relate together an anaphorically-referred-to chunk of preceding text with information presented by the speaker as 'not-at-issue', reversed pseudo-clefts are particularly suited to marking the conclusion of stages in the schematic structure of discourses. Along with clichés, generalizations, explicit repetitions and various other informationally-low forms, reversed pseudo-clefts serve appropriately as endings, lacking as they do the new information which can provide a basis for further discourse.)

* (There are three major informational types of cleft construction, one unmarked ('Type 1'), and two marked ('Type 2' and 'Type 3'). Unmarked clefts display a correlation between the structurally highlighted element and the locus of new information ('new' commonly in the sense of 'newly identified' or 'contrastive', rather than 'fresh'). (A characteristic of unmarked clefts is the very low communicative dynamism of information in the relative clause (which is often so evidently recoverable that it is ellipsed). It is common for the contrastive function of the highlighted element to go hand in hand with an explicit formulation of contrast.)

(In one type of marked cleft (Type 2) the highlighted element is typically short, and anaphoric or deictic. The relative clause is considerably longer, and contains new information. Type 3 clefts contain new (or, more accurately, 'semi-new') material in the highlighted element, commonly a circumstantial, or 'scene-setting', adjunct. They are classed as marked because, even though the highlighted element is non-given, it is the relative clause that is mainly responsible for conveying the message. Type 3 clefts are more common in LOB than LL, and are particularly favoured in formal, learned genres.)

The distribution of cleft and pseudo-cleft constructions, it has been demonstrated, is uneven across speech (in which pseudo-clefts are over three times more popular than clefts) and writing (in which clefts are marginally more popular). Reversed pseudo-clefts are particularly favoured in informal face-to-face dialogue (in the fictional categories of LOB they are almost entirely restricted to passages of dialogue). Clefts are more positively influenced than are pseudo-clefts in general by such factors as formality and degree of premeditation. (Within the informative categories of LOB, where clefts are almost twice as common as pseudo-clefts, a factor that strongly determines the frequency of clefts is the extent to which the discourse is opinionative and/or persuasive (as opposed to factual and/or descriptive).)

Several explanations may be advanced for the differences noted in the distribution between pseudo-clefts and clefts, mainly in terms of informational and thematic differences – but also certain formal and processing differences – between the constructions. (The popularity of pseudo-clefts (especially reversed) in speech is attributable largely to their givenness-orientation. Basic pseudo-clefts, we have seen, attach special status to background material, presenting it in the form of a subordinate clause which not only embodies a presupposition at the logico-semantic level, but also represents the theme at the textual level. It is not unduly surprising that a construction which so explicitly represents the background knowledge which the addressee is expected to be aware of, should occur more frequently in speech than writing. The basic pseudo-cleft functions then as an interpersonal 'tracking' device within the flow of discourse. The thematic, as well as informational, properties of basic pseudo-clefts are more suited to speech than writing: their thematic flexibility is of a type ('syntagmatic') that is germane to the dynamism of speech.) A language-processing interpretation of the structure of basic pseudo-clefts provides a further source of explanation for their popularity in

speech. The linear progression from explicitly represented non-news, to news, with the latter often very underdetermined by the former, offers the speaker an extended opportunity to formulate the message. Finally, a syntactic property of the basic pseudo-cleft construction that may be relevant to this discussion is its ready acceptance of a finite clause as highlighted element (where this is an uncommon choice with clefts, and does not occur, at least in LL or LOB, with reversed pseudo-clefts). This option is favoured in speech, with its tendency towards greater complexity in clausal interdependency.

The informational/thematic properties of reversed pseudo-clefts explain their popularity in speech (particularly in informal conversation between friends). The internal-referencing function of the construction, along with its generally low informativity, is well suited to the dynamic organization of spoken language. The typical realization of theme as a text-referential demonstrative enables a stretch of prior discourse (whose extent is likely to be larger with interlocutors whose acquaintance enables them to share a pool of common knowledge) to be identified with low-communicatively-dynamic information in the rheme/relative clause. Furthermore the almost cliché quality of many reversed pseudo-clefts, deriving from their minimal newsworthiness, is apposite in informal spoken genres. From a language-processing point of view, it may be noted that the theme of reversed pseudo-clefts may be referentially quite vague, and typically determines the rhematic content of the construction even less than is the case with the basic construction.

Cleft constructions exhibit several properties that account for their comparative popularity in written discourse (and particularly 'rhetorical' genres). By contrast with pseudo-clefts, their orientation is towards newness. New information is highlighted, via thematic prefiguration, both in unmarked clefts and in marked Type 2 clefts. Even though imbued with a non-controversial flavour, the new information in the relative clause of marked constructions is considerably higher in communicative dynamism than that of reversed pseudo-clefts. The denser information-packing of writing therefore provides one form of explanation for the register distribution of clefts. A further possible explanation relates to the lack of stress-marking in writing. Clefts may be used by the writer as a means of directing the reader into a particular interpretation of the information structure (namely one where the locus of new information is mapped on to the theme). It is precisely in 'opinionative' texts, where clefts occur with the greatest frequency, that one would expect writers to find a need for linguistic

means whereby intonation might be indicated, and in these that the contrastive implication generated by the theme/new combination would prove an attractive resource. A third reason for the comparative popularity of clefts in writing may be their structural similarity to impersonal constructions (such as 'It is said that . . .', 'It is true that . . .'), from which they derive a depersonalized quality and a formality that is often out of place in casual spoken genres. Fourthly, the 'paradigmatic' thematic flexibility of the cleft construction, which readily permits prepositional phrases and zero as the highlighted element, provides an attractive resource for writing, which arguably depends for its texture more on thematic than informational patternings. Finally, the preference for clefts in writing has a language-processing explanation. The presentation of news in the theme (at least in Types 1 and 3) necessitates pre-formulation of much of the content of the construction before it is encoded, a requirement that writers will be better able to manage, *ceteris paribus*, than speakers.

In this book I have adopted, and attempted to demonstrate the validity of, a comprehensive approach to the study of selected grammatical structures. By transcending the limited scope of a purely formal approach, and for that matter that of a solely discourse-based approach, it has been possible to pursue integrated explanations for the characteristics of the cleft and pseudo-cleft constructions. There is still a need for research to be conducted in a number of areas that have been touched upon, but not explored extensively, in the present study. The formal semantics of clefts and pseudo-clefts would repay further investigation, and particularly the complexities of their token-value structure. Despite the considerable amount of work that has been done – largely within a transformational framework – on the formal syntax of the constructions, there is still a shortage of convincing analyses. A wide-ranging comparison of the syntactic properties of clefts and pseudo-clefts with those of agnate constructions would provide a clearer picture of their distinctive formal features.

There is still much to be done on the textual functions of language, and this will must certainly be facilitated by the availability in the future of computer corpora many times the size of those used in this study. It will be of interest, for instance, to see if any of the tentative suggestions that have been made regarding differences in the textual organization of speech and writing (that the primary form of organization in the former is given-new structure, and theme-rheme structure in the latter), are borne out, or disconfirmed, by subsequent

investigations. One valuable lead that has been provided for further investigation of the discourse functions of particular constructions, is Fries's study of method of development through theme-rheme selection. Text-based research of this kind would benefit from a statistical approach, supported by notions of linguistic probability.

Notes

1 APPROACHES TO THE STUDY OF CLEFT AND PSEUDO-CLEFT CONSTRUCTIONS

- 1 Halliday (1970a, 1985) uses the term to refer to the 'climax' of new information within the information unit.
- 2 The term is suggested by Huddleston (1984:459). It refers to the constituent which, in propositional semantic terms, has the role of 'identifier', which in syntactic terms has the role of 'complement of *be*' (sometimes called the 'post-copular constituent'), which in textual semantic terms has the role – in unmarked cases – of 'new' or 'comment', and which in logical terms has the role of 'focus'. The term 'highlighted element' satisfies the need for an expression which is neutral as to the semantic/syntactic/textual/logical role of the constituent in question.
- 3 The terms 'identifying' and 'attributive' are used by Halliday (1985). They correspond to 'specificational' and 'predicational' as used by a number of transformational linguists, including Akmajian (1979) and Higgins (1979). Other terms that have been used include Kuro and Wongkhamhong's (1981) 'identificational' and 'characterizational' and, for 'identifying', 'equational' (Huddleston 1971) and 'equative' (Bolinger 1972a).
- 4 These include Akmajian 1970; Chomsky 1970; Delahunty 1982; Gundel 1977; Halvorsen 1978; Higgins 1979; Peters and Bach 1971; Pinkham and Hankamer 1975.
- 5 See Halvorsen (1978), Delahunty (1982).
- 6 See Jespersen (1927) for his first analysis, and Jespersen (1937) for a second.
- 7 See Daneš (1964), Firbas (1964), Vachek (1966).

2 THE DATABASE

- 1 See for example Halliday (1967a, 1985), Gundel (1977), and Prince (1978).
- 2 Prince (1978), Jones and Jones (1985), and Sornicola (1988) are text-based studies.

- 3 The recordings were made without prior knowledge of the main participants, and to prevent the possibility of identification, the names used in the transcriptions are fictitious but prosodically equivalent to the originals. The contributions of non-surprised speakers (participants who had knowledge of the recording and whose task it was to keep the conversation going) were not included – in the case of Texts S.1–S.3 and S.7.1, but were included – in the case of the remaining texts – in the total 5,000 words of each text. The utterances of non-surprised speakers were not prosodically analysed.
- 4 Explanations of the prosodic system are to be found in Crystal (1969) and Quirk *et al.* (1985: Appendix II).
- 5 See further Gregory and Carroll (1978:7–8).
- 6 See, for example, O'Donnell (1974), Olson (1977), Chafe (1982).

3 DEFINING THE CLASS

- 1 The term 'fused relative' is taken from Huddleston (1984:402). Quirk *et al.* (1985: §15.8) use 'nominal relative clause'; the term 'free relative' is widely used amongst transformational linguists.
- 2 The term 'wh-cleft' is used by Hankamer (1974) and Prince (1978). Prince refers to a paper read by Catherine N. Ball at the 1977 University of Pennsylvania Linguistics Colloquium, entitled 'Th-clefts'.
- 3 The material in Section 3.1.1 has been adapted from Collins (1985).
- 4 For information on the transcription symbols used in the LL extracts see the list of abbreviations and symbols on pp.xiv–xv.
- 5 As noted in Collins (1982) it is arguably a relatively superficial fact of English that modals cannot co-occur (resulting from the fact that they have no non-finite forms). It is possible to have two expressions of modality in a single clause (for example 'Very seldom will it be possible for good to be done in this field under six months').
- 6 For a discussion of restrictions on clefts with highlighted adjective phrase, see Section 4.3.
- 7 Declerck (1983b:42) defines a determinative pronoun as 'a pronoun that is restrictively modified by a relative clause, participle or prepositional phrase'. Another possibility is that the *it* of proverbial sentences is the subject slot filler of a dislocated structure (see Poutsma 1928:992, Jacobsson 1971:321).
- 8 The examples are borrowed from Declerck (1983b:15), who in turn borrowed them from Poutsma (1928:730) and Visser (1970:50).
- 9 In fact there are no tokens of the type in (51) in the corpus. There are, on the other hand, five tokens of the type in (49) and (50) (including these).
- 10 All the latter occur in the context of casual dialogue, actual or represented.
- 11 The term 'circumstantial attribution' is borrowed from Halliday (1985:112ff).
- 12 There were only four occurrences of clefts like those in (52) in the corpus, all from LOB. Curiously, Huddleston (1984:460) queries the acceptability of sentences like (52): 'The *wh* word cannot, or can scarcely, follow a preposition or longer sequence: we would use (23iii) [*It was Ed that she*

- was referring to"] or (iv) [*It was to Ed that she was referring*] rather than [*It was Ed to whom she was referring*].
- 12 Jespersen observes that Irish English makes 'excessive use of cleft sentences' and gives several examples from Irish literature of sentences similar to the two examples, including *It is angry that he was* and *It's an angel you are to forgive me*.
- 13 See further Plötz (1972:37).

4 FORMAL PROPERTIES

- 1 The term 'pseudo-cleft' appears to be an invention of the transformational school. To the knowledge of the present author, it does not appear in any linguistic publication prior to the advent of that school.
- 2 Simnkal (1982) notes that there are four constructions which permit 'subject relative pronoun absence'. As well as clefts, there are 'matrix sentence constituents' of the following type:

- (i) *there* + *be* + NP (for example *There's a girl wants to see you.*)
- (ii) NP + *have* + NP (for example *I had a sister married an Italian.*)
- (iii) NP + *be* + NP (for example *I'm the only guy knows how to fix it.*)

Simnkal argues that all four constructions have a common underlying semantic representation: the matrix sentence 'postulates the existence of an entity which is then qualified in some way by the relative clause that follows' (p.324). The effect is to foreground a noun phrase which then becomes the focus of the whole sentence, and which 'is often assigned more prominent stress than the other constituents of the sentence' (p.324). Interestingly, in the spoken data, 'subject relative pronoun absence' occurred almost exclusively in clefts with a focal highlighted item (that is, in cases of the type 'It was TOM offered it to SUE'; rather than 'It was Tom offered it to SUE').

- 3 The term 'embedded' has a variety of senses in modern grammars. In some grammars, all cases of clause subordination are handled in terms of embedding. Halliday (1985), and members of the Systemic-Functional School, restrict embedding to non-immediate clause constituents. Embedded, or 'rankshifted', clauses are those which function within the structure of a group. Huddleston's interpretation of the term is intermediate between these broader and narrower senses. He argues that there are cases of clause subordination that are not to be handled in terms of embedding. Thus, in *Ed liked it, whereas Max thought it appalling* (= Huddleston (3), p.379), *whereas Max thought it appalling* is subordinate, but not embedded in the superordinate clause, *Ed liked it*; rather, Huddleston suggests, the subordinate clause is an immediate constituent of the sentence. Thus Huddleston's class of embedded clauses includes some which in Halliday's model would be classified as 'hypotactic'.
- 4 See for example Akmajian (1970).
- 5 It is not entirely clear to me why this should be so. The fact that reversed pseudo-clefts with highlighted *do*-complement are not ungrammatical (*Wash the dishes was what she did next*; compare **It was wash the dishes*

that she did next) would seem to remove the possibility of a thematic explanation. Perhaps the explanation lies in the different types of relative clause that occur in pseudo-clefts.

- 6 Halliday's (1967a:227-31) seminal discussion of the token-value distinction is introduced with the warning that 'the discussion is extremely tentative; much more study is required of equative clauses as such'. The complexities of Halliday's account have baffled some. For example Higgins (1979), in his comprehensive study of pseudo-clefts, makes the admission that 'I have been unable, in spite of much trying, to make clear to myself the distinctions which Halliday draws (1967,69-70;227-31) between the two oppositions "value" and "variable", and "identifier" and "identified", or to find all the ambiguities which his analysis predicts, and therefore I have abandoned his approach' (p.277).
- 7 See Grice (1975), Karttunen and Peters (1975), and Harnish (1976).
- 8 For example Chomsky (1970), Prince (1978).

5 COMMUNICATIVE MEANINGS

- 1 The answer is 'When the water is turned into ice.'
- 2 This use of the term 'antecedent' is borrowed from Clark and Haviland (1977) who, using as a basis Grice's (1975) 'cooperative principle', formulate a 'maxim of antecedence' to account for the cooperative speaker's attempt to present information in such a way that the addressee's job of decoding is maximally facilitated (p.4): 'Maxim of Antecedence: Try to construct your utterance such that the listener has one and only direct antecedent for any given information and that it is the intended antecedent.'
- 3 There is clearly an inconsistency between this definition of 'knownness', which focuses on the hearer's general knowledge store, and that in Prince (1978) which focuses on the knowledge store of people other than the hearer. Since it would be difficult in practice to distinguish between the knowledge of the hearer and other people, I shall henceforth assume the Prince (1981) definition when referring to 'known' information.
- 4 It should be noted that the relative clause of *it*-clefts is not a noun phrase.

6 COMMUNICATIVE MEANINGS IN THE CORPUS

- 1 It is probably preferable to extend Halliday's notion of 'contrastive' in this way, in order to maintain the theoretically significant generalization that nucleus-bearing items represent new information, rather than to abandon this generalization as a means of accounting for recalcitrant data. See, for example, Taglicht's (1984:42) suggestion that the relationship between 'new information' and focus-placement (marked, in his proposed model, by non-nuclear as well as nuclear accents) is as follows: 'All "new information" is represented by focal items, but not all focal items represent "new information".'
- 2 Taglicht (1984:40) observes that when 'operators' (in the sense of Quirk *et al.* 1985:2.48) bear accents, 'it is usually not the grammatical function of the individual operator that is marked as focal, but rather the "invisible" component that is common to the whole class, viz. the "assertive"

component of the operator.' It would appear from examples such as (7) that the 'invisible' component of 'assertiveness' is not restricted to operators, but lies dormant in any finite verb, capable of being brought forth if the verb carries stress in certain appropriate contexts.

7 CLEFTS, PSEUDO-CLEFTS, AND REGISTER VARIATION

- 1 Frequencies in Table 7.3 have been calculated with respect to the total number of words in each corpus. It must be admitted that a more revealing standard for comparison might have been the number of clauses, rather than the total running words (see discussion of lexical density in Section 2.3). Unfortunately, however, no syntactically parsed versions of L.OB or L.L are yet available, and the task of manually processing the two corpora for clause numbers would have been too time-consuming to be feasible in the present study.
- 2 The metaphor of a snowball for the spoken language is borrowed from Mitchell (1975:186).
- 3 Unfortunately, no comparisons can be made between Biber's and the present study with respect to pseudo-clefts, because these were dropped from Biber's analysis as a feature that emerged from the factor analysis lacking salient weight (see Biber 1986:393).
- 4 Michael Halliday has suggested (personal communication) that the preference for indirect antecedence in private spoken texts may be due to the greater potential for 'subtlety' (compare humour) in private than public contexts.
- 5 Some striking word-length differences emerge if we compute average clause lengths for the three examples analysed in Halliday's (1987) Table 1. For Extract 1A, a recording of spontaneous spoken discourse, the average clause length is 6.4 words; for Extract 1B, an informal written version, the average is 8.5 words; and for Extract 1C, a formal written version, it is 13.8 words.

Bibliography

- Abercrombie, D. 1963. Conversation and spoken prose. *English Language Teaching* 18.1: 10-16.
- Abercrombie, D. 1971. *Studies in Phonetics and Linguistics*. London: Oxford University Press.
- Akmajian, A. 1970. On deriving cleft sentences from pseudo-cleft sentences. *Linguistic Inquiry* 1: 149-68.
- Akmajian, A. 1979. *Aspects of the Grammar of Focus in English*. New York: Garland Publishing.
- Allerton, D.J. 1978. The notion of 'givenness' and its relations to presupposition and to theme. *Lingua* 44: 133-68.
- Atlas, J.D. and Levinson, S.C. 1981. *It-clefts, informativeness, and logical form: radical pragmatics* (Revised Standard Version). In P. Cole (ed.) 1981. *Radical Pragmatics*. New York: Academic Press. 1-61.
- Beaugrande, R. de and Dressler, W. 1981. *Introduction to Text Linguistics*. London: Longman.
- Biber, D. 1985. Investigating macroscopic textual variation through multi-lecture/multi-dimensional analyses. *Linguistics* 23.2: 337-60.
- Biber, D. 1986. Spoken and written textual dimensions in English: resolving the contradictory findings. *Language* 62.2: 384-414.
- Bolinger, D. 1972a. A look at equatives and cleft sentences. In E. Fichow (ed.) 1972. *Studies for Einar Haugen*. The Hague: Mouton. 96-114.
- Bolinger, D. 1972b. *That's that*. The Hague: Mouton.
- Bolinger, D. 1972c. Accent is predictable (if you're a mind-reader). *Language* 48: 633-44.
- Brazill, D.C., Coulthard, M. and Johns, C. 1980. *Discourse Intonation and Language Teaching*. London: Longman.
- Bresnan, J.W. 1971. Sentence stress and syntactic transformations. *Language* 47: 257-81.
- Brown, G., Currie, K. and Kenworthy, J. 1980. *Questions of Intonation*. London: Croom Helm.
- Brown, G. and Yule, G. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.
- Chafe, W. 1970. *Meaning and the Structure of Language*. Chicago: University of Chicago Press.
- Chafe, W. 1974. Language and consciousness. *Language* 50: 111-33.
- Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li (ed.) 1976. *Subject and Topic*. New York: Academic Press. 25-56.
- Chafe, W. 1982. Integration and involvement in speaking, writing, and oral literature. In D. Tannen (ed.) 1982. *Spoken and Written Language: Exploring Orality and Literacy*. Norwood, NJ: Ablex. 35-54.
- Chomsky, N. 1970. Remarks on nominalization. In R.A. Jacobs and P.S. Rosenbaum (eds) 1970. *Readings in English Transformational Grammar*. Waltham, Mass.: Ginn. 184-221.
- Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English*. New York: Harper & Row.
- Clark, H.H. and Haviland, S.E. 1977. Comprehension and the given-new contract. In R. Freedle (ed.) 1977. *Discourse Production and Comprehension*. Hillsdale, NJ: Erlbaum. 1-40.
- Cole, P. (ed.) 1981. *Radical Pragmatics*. New York: Academic Press.
- Collins, P.C. 1982. Cleft sentences in English discourse. *Australian Review of Applied Linguistics* 5.1: 60-83.
- Collins, P.C. 1985. *Th-clefts and all-clefts*. In J.E. Clark (ed.) 1985. *The Cultivated Australian: Festschrift in Honour of Arthur Delbridge*. Hamburg: Buske. 45-53.
- Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.
- Crystal, D. 1975. *The English Tone of Voice: Essays in Intonation, Prosody and Paralinguistics*. London: Edward Arnold.
- Curme, G.O. 1931. *A Grammar of the English Language* vol. 3. Boston: D.C. Heath & Co.
- Dahl, Ö. What is new information? In N.E. Enkvist and V. Kohonen. 1976. *Reports on Text Linguistics: Approaches in Word Order*. Meddelanden från Stiftelsen för Åbo Akademi Forskningsinstitut no.8. Åbo/Turku. 37-50.
- Daneš, F. 1964. A three-level approach to syntax. *Travaux linguistiques de Prague* 1: 225-40.
- Declerck, R. 1981. Pseudo-modifiers. *Lingua* 54: 135-63.
- Declerck, R. 1983a. 'It is Mr Y' or 'He is Mr Y'? *Lingua* 59: 209-46.
- Declerck, R. 1983b. Predicational clefts. *Lingua* 61: 9-45.
- Declerck, R. 1988. *Studies on Copular Sentences, Clefts and Pseudo-clefts*. Leuven: Leuven University Press.
- Delahunty, G. 1982. *Topics in the Syntax and Semantics of English Cleft Sentences*. Bloomington, Indiana: Indiana University Linguistics Club.
- Delahunty, G. 1984. The analysis of English cleft sentences. *Linguistic Analysis* 13: 63-113.
- Emonds, J. 1976. *A Transformational Approach to English Syntax*. New York: Academic Press.
- Firbas, J. 1957. K otázce nezákladových podmětů v současné angličtině [On the problem of non-thematic subjects in contemporary English]. *Časopis pro Moderní Filologii* 39: 22-42. 165-73.
- Firbas, J. 1964. On defining the theme in functional sentence perspective. *Travaux linguistiques de Prague* 1: 267-80.
- Firbas, J. 1979. First intonation centre prosodic shade in the modern English clause. In S. Greenbaum, G. Leech and J. Svartvik (eds) 1979. *Studies in*

- English Linguistics for Randolph Quirk*. London: Longman. 125-33.
- Fries, P.H. 1981. On the status of theme in English: arguments from discourse. *Forum Linguisticum* 6.1: 1-38.
- Fronek, J. 1983. Some criticisms of Halliday's 'information systems'. *Lingua* 60: 311-29.
- Gilhotra, M. 1984. On the nature of presuppositions. *Newcastle Working Papers in Linguistics* 1: 36-68.
- Givón, T. (ed.) 1983. *Topic Continuity in Discourse: A Quantitative Cross-language Study* (Typological Studies in Language, 3). Amsterdam/Philadelphia: John Benjamins.
- Green, G.M. 1971. Notes on clefts and pseudo-clefts and related matters. *Studies in the Linguistic Sciences* 1: 1-7.
- Gregory, M.J. and Carroll, S. 1978. *Language and Situation: Language Varieties and Their Social Contexts*. London: Routledge & Kegan Paul.
- Grice, H.P. 1975. Logic and conversation. In P. Cole and J.L. Morgan (eds) 1975. *Studies in Syntax 3: Speech Acts*. New York: Seminar Press. 41-58.
- Gundel, J.K. 1977. Where do cleft sentences come from? *Language* 53: 543-59.
- Gundel, J.K. 1985. 'Shared knowledge' and topicality. *Journal of Pragmatics* 9: 83-107.
- Halliday, M.A.K. 1967a. Notes on transitivity and theme in English, Part 2. *Journal of Linguistics* 3: 199-244.
- Halliday, M.A.K. 1967b. *Intonation and Grammar in British English*. The Hague: Mouton.
- Halliday, M.A.K. 1968. Notes on transitivity and theme in English, Part 3. *Journal of Linguistics* 4: 153-308.
- Halliday, M.A.K. 1970a. *A Course in Spoken English: Imposition*. London: Oxford University Press.
- Halliday, M.A.K. 1970b. Language structure and language function. In J. Lyons (ed.) 1970. *New Horizons in Linguistics*. Harmondsworth: Penguin. 140-65.
- Halliday, M.A.K. 1970c. Functional diversity in language as seen from a consideration of modality and mood in English. *Foundations of Language* 6: 322-61.
- Halliday, M.A.K. 1973. *Explorations in the Functions of Language*. London: Edward Arnold.
- Halliday, M.A.K. 1975. Language as social semiotic: towards a general sociolinguistic theory. In A. Makkai and V.B. Makkai (eds) 1975. *The First LACUS Forum*. Columbia, SC: Hornbeam. 17-46.
- Halliday, M.A.K. 1977. Text as semantic choice in social contexts. In T.A. van Dijk and J. Petöfi (eds) 1977. *Grammars and Descriptions*. Berlin and New York: de Gruyter. 176-225.
- Halliday, M.A.K. 1978. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Halliday, M.A.K. 1979. Differences between spoken and written language: some implications for literacy teaching. In G. Page, J. Elkins and B. O'Connor (eds) 1979. *Communication through Reading* (vol. 2). Adelaide: Australian Reading Association. 37-52.
- Halliday, M.A.K. 1985. *An Introduction to Functional Grammar*. London: Edward Arnold.

- Halliday, M.A.K. 1987. Spoken and written modes of meaning. In R. Horowitz and J. Samuels (eds) 1987. *Comprehending Oral and Written Language*. New York: Academic Press. 55-82.
- Halliday, M.A.K. The meaning of Modern English. Unpublished manuscript.
- Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English*. London: Longman.
- Halvorsen, P.-K. 1978. The syntax and semantics of cleft constructions. *Texas Linguistic Forum, II*. Department of Linguistics, University of Texas. Austin.
- Hankamci, J. 1974. On the non-cyclic nature of WH-clefting. *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*. University of Chicago: Chicago Linguistic Society. 221-32.
- Harnish, R.M. 1976. Logical form and implicature. In T.G. Bever, J.J. Katz and D.T. Langendoen (eds) 1976. *An Integrated Theory of Linguistic Ability*. New York: Crowell. 313-91.
- Harries-Delisle, J. 1978. Contrastive emphasis and cleft sentences. In J.H. Greenberg (ed.) 1978. *Universals of Human Language: Vol. 4. Syntax*. Stanford: Stanford University Press. 420-86.
- Haviland, S.E. and Clark, H.H. 1974. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behaviour* 13: 512-21.
- Higgins, F.R. 1979. *The Pseudo-cleft Construction in English*. New York: Garland Publishing.
- Hornby, P.A. 1971. The role of topic-comment in the recall of cleft and pseudo-cleft sentences. *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society*. University of Chicago: Chicago Linguistic Society. 445-53.
- Huddleston, R.D. 1971. *The Sentence in Written English*. Cambridge: Cambridge University Press.
- Huddleston, R.D. 1984. *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.
- Jacobson, B. 1971. A note on the classification of relative nominal clauses. *Moderna Språk* 65: 312-22.
- Jespersen, O. 1927. *A Modern English Grammar on Historical Principles. Part 3: Syntax*. London: Allen & Unwin.
- Jespersen, O. 1937. *Analytic Syntax*. London: Allen & Unwin.
- Jespersen, O. 1958. *A Modern English Grammar on Historical Principles. Part 7: Syntax*. London: Allen & Unwin.
- Johansson, S., Leech, G. and Goodluck, H. 1978. *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Oslo: Department of English, University of Oslo.
- Jones, L.B. and Jones I.K. 1985. Discourse functions of five English sentence types. *Word* 36.1: 1-21.
- Joos, M. 1967. *The Five Clocks*. New York: Harcourt, Brace & World.
- Karttunen, L. and Peters, S. 1975. Conventional implicature in Montague Grammar. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*. University of California: Berkeley Linguistics Society.
- Karttunen, L. and Peters, S. 1979. Conversational implicature. In C.-K. Oh and D. Dinneen (eds) 1979. *Syntax and Semantics II: Presupposition*. New

- York: Academic Press. 1-56.
- Keenan, E. 1971. Two kinds of presupposition in natural language. In C. Fillmore and D.T. Langendoen (eds) 1971. *Studies in Linguistic Semantics*. New York: Holt. 45-54.
- Kuno, S. 1972. Functional sentence perspective. *Linguistic Inquiry* 3: 269-320.
- Kuno, S. 1975. Conditions for verb-phrase deletion. *Foundations of Language* 13: 161-75.
- Kuno, S. 1978. Generative discourse analysis in America. In W. Dressler (ed.) 1978. *Current Trends in Textlinguistics*. Berlin and New York: de Gruyter. 275-94.
- Kuno, S. 1979. On the interaction between syntactic rules and discourse principles. Unpublished manuscript.
- Kuno, S. and Wongkhamthong, P. 1981. Characterizational and identification sentences in Thai. *Studies in Language* 5: 65-109.
- Lees, R.B. 1963. Analysis of the 'cleft sentence' in English. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 16: 371-88.
- Martin, J.R. 1984a. Language, register and genre. In F. Christie (ed.) 1984. *Language Studies: Children Writing*. Geelong, Vic.: Deakin University Press. 21-30.
- Martin, J.R. 1984b. The development of register. In J. Fiac and R.O. Freedle 1984. *Developmental Issues in Discourse*. Norwood, NJ: Ablex. 1-39.
- Martin, J.R. Forthcoming. *English Text: System and Structure*. Oxford: Frances Pinter.
- Mathesius, V. 1939. O tak zvaném aktuálním členění větném. [On the so-called functional sentence perspective] *Slovo a Slovesnost* 5: 171-4.
- Michell, T.F. 1975. *Principles of Firthian Linguistics*. London: Longman.
- Morgan, J.L. 1975. Sentence fragments and the notion 'sentence'. In Kachru et al. (eds) 1973. *Issues in Linguistics: Papers in Honor of Henry and Renée Kahane*. Urbana: University of Illinois Press. 719-57.
- O'Donnell, R.C. 1974. Syntactic differences between speech and writing. *American Speech* 49: 102-10.
- Olson, D.R. 1977. From utterance to text: the bias of language in speech and writing. *Harvard Educational Review* 47: 257-81.
- Peters, S. and Bach, E. 1971. Pseudo-cleft sentences. In *Report to NSF: On the Theory of Transformational Grammar*. GS-2468. Department of Linguistics, University of Texas at Austin.
- Pinkham, J. and Hankamer, J. 1975. Deep and shallow clefts. *Papers from the Eleventh Regional Meeting of the Chicago Linguistic Society*. University of Chicago: Chicago Linguistic Society. 429-50.
- Plotz, S. 1972. *Simple Copula Sentences in English*. Frankfurt/M.: Athenäum Verlag.
- Poutsma, H. 1916. *A Grammar of Late Modern English. Part 2: The Parts of Speech. Section I*. Groningen: P. Noordhoff.
- Poutsma, H. 1928. *A Grammar of Late Modern English. Part 1: The Sentence. First Half. The Elements of the Sentence*. Groningen: P. Noordhoff.
- Prince, E.F. 1978. A comparison of WH-clefts and *it*-clefts in discourse. *Language* 54: 883-906.

- Prince, E.F. 1979. On the given/new distinction. *Papers from the Fifteenth Regional Meeting of the Chicago Linguistic Society*. University of Chicago: Chicago Linguistic Society. 362-76.
- Prince, E.F. 1981. Toward a taxonomy of given-new information. In P. Cole (ed.) 1981. *Radical Pragmatics*. New York: Academic Press. 223-55.
- Prince, E.F. 1985. Fancy syntax and shared knowledge. *Journal of Pragmatics* 9: 65-81.
- Quirk, R. 1974. *The Linguist and the English Language*. London: Edward Arnold.
- Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1972. *A Grammar of Contemporary English*. London: Longman.
- Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Shankal, A. 1982. There's a lot mightn't believe this . . . : variable subject relative pronoun absence in Australian English. In D. Sankul (ed.) 1982. *Variation Omnibus* (Current Inquiry into Language, Linguistics and Human Communication 40). Edmonton: Linguistic Research. 321-8.
- Sannicola, R. 1988. *It*-clefts and *Wh*-clefts: two awkward sentence types. *Journal of Linguistics* 24: 343-79.
- Stubbs, M. 1983. *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. Oxford: Basil Blackwell.
- Svartvik, J. and Egg-Olofsson, M. 1982. Tagging the London-Lund Corpus of Spoken English. In S. Johansson (ed.) 1982. *Computer Corpora in English Language Research*. Bergen: Norwegian Computing Centre for the Humanities. 85-102.
- Svartvik, J. and Quirk, R. (eds) 1980. *A Corpus of English Conversation*. Lund: CWK Gleerup.
- Taglicht, J. 1984. *Message and Emphasis: On Focus and Scope in English*. London: Longman.
- Trávníček, F. 1962. O tak zvaném aktuálním členění větném. [On the so-called functional sentence perspective] *Slovo a Slovesnost* 22: 163-71.
- Ure, J. 1971. Lexical density and register differentiation. In G.E. Perren and J.L.M. Tilm (eds) 1971. *Applications of Linguistics: Selected Papers of the Second World Congress of Applied Linguistics, Cambridge 1969*. Cambridge University Press. 443-52.
- Ure, J. and Ellis, G. 1977. Register in descriptive linguistics and linguistic sociology. In O. Uribe-Villegas (ed.) 1977. *Issues in Sociolinguistics*. The Hague: Mouton. 197-243.
- Vachek, J. 1966. *The Linguistic School of Prague*. Bloomington, Ind.: Indiana University Press.
- van Dijk, T.A. 1977. Sentence topic and discourse topic. *Papers in Slavic Philology* 1.49-61.
- Visser, F.T. 1970. *An Historical Syntax of the English Language, Part 1: Syntactical Units with One Verb*. Leiden: E.J. Brill.

Index



Abercrombie, D. 21, 22
 Akmajian, A. 50, 54, 217, 219
all-clefts 32-4
 Allerton, D. 94, 95
 at-issueness 106-7
 Atlas, J. 73
 attribution 36-44

 Bach, F. 217
 Ball, C. 218
 base derivation 49, 50
 Beaugrande, R. de 91
 Biber, D. 18, 23-5, 182, 200, 221
 Bolinger, D. 85, 154, 217
 Brazil, D. 88
 Bresnan, J. 85
 Brown Corpus 11, 16, 17, 24
 Brown, G. 88, 95

 Carroll, S. 218
 Chafe, W. 4, 85, 86, 92, 94, 99,
 103-4, 218
 Chomsky, N. 3, 85, 217, 220
 Clark, H. 92, 99, 104, 220
 clefts; class of 34-6
 Collins, P. 203, 218
 communicative variation 4-5
 communicative dynamism 79, 94,
 107
 connectedness 53
 contrastiveness 87-8, 89-90, 110,
 141-3
 corpus-based approach 10-12, 208
 co-text versus context 96
 Crystal, D. 85, 218
 Curme, G. 47

Dahl, O. 94
 Daneš, F. 85, 217
 Declerck, R. 40-1, 47, 48, 100-2,
 209, 218
 Delahanty, G. 36, 55, 57, 73-4, 217
do 127-9
 Dressler, W. 91
dummy it 50

 Eeg-Olofsson, M. 14
 Ellis, G. 6
 Emonds, J. 50
end-weight 61
 entailment 73
 exclusiveness implicature 32-4,
 69-71
 extended text reference 140

 field 7, 16-18, 126-30
 Filbas, J. 77-9, 85, 86, 88-9, 217
 focus, 2, 43; marked 83, 101
 Fries, P. 77, 81, 86, 216
 Fronek, J. 88-9
 functional grammar 3-6
 functional sentence perspective 4, 5,
 76

 Gilhott, M. 70
 givenness 85-8, 89, 91-3; and
 presupposition 93-4; taxonomies
 of 94-7; in clefts and pseudo-
 clefts 98-116
 Givón, T. 81
 Goodluck, H. 16
 Green, G. 49
 Gregory, M. 218

Grice, H. 4, 70, 204, 220
 Gundel, J. 50, 97, 100, 102-5, 109,
 217

 Halle, M. 85
 Halliday, M. 2, 3-4, 5-6, 7, 22-3,
 29-30, 31, 33, 37, 42, 61, 65,
 68-9, 75-85, 86-91, 95, 103, 117,
 120, 127, 135, 139-40, 155, 156,
 170, 182, 200, 205, 217, 218, 219,
 220, 221
 Halvorsen, P.-K. 26, 49, 69, 70, 71,
 72-3, 210, 217
 Hankamer, J. 217, 218
happen 126-8
 Harnish, R. 220
 Harries-Delisle, H. 155, 208
 Hasan, R. 29-30, 95, 127, 140
 Haviland, S. 92, 220
 Higgins, F. 26, 49, 71, 208, 217, 220
 highlighted element 54-67
 Hornby, P. 208
 Huddleston, R. 49, 50, 52-3, 97,
 106-9, 112, 217, 218, 219

 identification 36-44, 67-9
 inferability 110
 information 6, 76, 85-116, and basic
 pseudo-clefts 121-32, 189-96; and
 clefts 158-70, 189-96
 intonation; and basic pseudo-clefts
 117-20; and clefts 154-8

 Jacobsson, B. 218
 Jespersen, O. 3, 10, 39-40, 47, 48,
 217
 Johansson, S. 16
 Jones, L.B. 217
 Jones, L.K. 217
 Joos, M. 19

 Karttunen, L. 69, 70, 220
 Keenan, E. 72
 Knoo, S. 91, 112, 217
 knownness 106, 109
 Kruisinga, E. 10

 Lancaster-Oslo/Bergen Corpus
 16-25
 Leech, G. 16

Lees, R. 50
 Levinson, S. 73, 208
 lexical density 22
 London-Lund Corpus 12-25

 Martin, J. 18-19, 19-20, 81
 Mathesius, V. 77-8
 Mitchell, T. 221
 mode 7, 19-21, 131-2
 Morgan, J. 49

 O'Donnell, R. 218
 Olson, D. 218

 Peters, S. 69, 70, 217, 220
 Pinkham, J. 217
 Plötz, S. 219
 postponed-modifier analysis 50-2
 Poutsma, H. 10, 47-8, 218
 Prague School, S. 77, 85, 86, 94
 predication clefts 41-2, 44
 presupposition: existential 71-4;
 and givenness 93-4
 Prince, E. 3, 4, 26, 34, 36, 57, 58,
 91, 92, 94-5, 97-100, 101, 103-4,
 105, 109, 110, 112, 130, 187-9,
 195, 202-4, 208, 210, 217, 218,
 220
 prosody 14-15
 proverbial sentences 40-1
 pseudo-clefts; class of 26-34

 Onirk, R. 10-11, 12, 49, 52, 61,
 155, 218, 220

 register 6-8, 16-25, 178-207; and
 highlighted elements 196-202
 relative clause 2, 52

 shared knowledge 92
 Shukal, A. 219
 Sornicola, R. 217
 speech, 182-5; versus writing 21-5,
 178-82
 Stnbs, M. 145-6
 subordination: non-embedded 52-3
 Survey of English Usage 11-12
 Svartvik, J. 10, 14
 system deviance 44-8

 Taglicht, J. 42-3, 220

230 *Index*

- tenor 7, 18-19, 130-1, 183
 textual component 75-7
th-clefts 29-31
 theme 77-85, and basic pseudo-
 clefts 132-8; and reversed
 pseudo-clefts 151-4; and clefts
 170-7
 tone unit 86
 topic 102, 105
 total accountability 44, 208
 transitivity 5-6
 Trávníček, F. 77, 78, 86
 Ure, J. 6, 22
 Vachek, J. 217
 van Dijk, T. 79
 Visser, F. 218
wh-clefts 29-31
 Whorf, B. 188
wh-words 34
 Wongkhomthong, P. 217
 writing 186-9; versus speech 21-5,
 178-82
 Yule, G. 95

